



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 175929**

**TO: Manjunath N Rao**  
**Location: REM-2A01/2C70**  
**Art Unit: 1652**  
**Tuesday, January 17, 2006**  
**Case Serial Number: 10/618976**

**From: Kristine Hensle**  
**Location: Biotech-Chem Library**  
**REM-1B69**  
**Phone: (571) 272-4161**

**Kristine. Hensle@uspto.gov**

### **Search Notes**

Examiner Rao,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Kristine Hensle  
Librarian  
STIC Biotech/Chem Library  
(571)272-4161

175929

**STIC-Biotech/ChemLib**

---

**From:** Rao, Manjunath N.  
**Sent:** Saturday, January 07, 2006 1:21 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** Sequence search request for 10619359

**From:** Manjunath N. Rao  
**Art Unit** 1652, Room 2A01  
**Mail Box in Room** 2C70  
**Phone:** 272-0939

**Date:** 1-7-06

Please search the following as soon as possible for application with serial number  
**10/618976**

1. **SEQ ID NO: 1** against all commercial nucleic acid sequence databases, issued patents/published applications nucleic acid sequence database and pending application nucleic acid sequence database. Please provide a print of results
2. **SEQ ID NO: 2** against all commercial amino acid sequence databases, issued patents/published applications amino acid sequence database and pending application amino acid sequence database. Please provide a print of results.

If you have any questions please call me at the above phone number.

Thanks



# STIC SEARCH RESULTS FEEDBACK FORM

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or contact:

Mary Hale, Information Branch Supervisor  
Remsen Bldg. 01 D86  
571-272-2507

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC Biotech-Chem Library Remsen Bldg



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2006, 10:11:00 ; Search time 7422 Seconds  
(without alignments)  
11511.144 Million cell updates/sec

Title: US-10-618-976-1

Perfect score: 1503

Sequence: 1 atgcaactcaacgcgctttt.....ccgagccagccagcgct 1503

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.in.\*

3: gb.env.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pr.\*

9: gb.ro.\*

10: gb.sts.\*

11: gb.sy.\*

12: gb.un.\*

13: gb.vi.\*

14: gb.htg.\*

15: gb.pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1503	100.0	1503	6	AR393878 Sequence
2	647.4	43.1	1503	1	AY59352 Inermus a
3	638.4	42.5	1503	6	E30186 Thermophil
4	638.4	42.5	2085	1	AB016244 Thermus a
5	638.4	42.5	273285	1	AE017304 Thermus t
6	636.8	42.4	1500	6	BD179859 Highly th
7	636.8	42.4	1503	6	AX801458 Sequence
8	636.8	42.4	110000	1	Continuation (12 o
9	636.8	42.4	110000	1	Continuation (13 o
10	596.2	39.7	1503	1	AY459351 Thermus s
11	492.8	32.8	10747	1	AE002007 Deinococc
12	421.8	28.1	110000	1	Continuation (13 o
13	415.6	27.7	110000	1	Continuation (36 o
14	403.2	26.8	110000	1	Continuation (6 of
15	329.2	21.9	110000	1	Continuation (3 of
16	326.4	21.7	307050	1	BX294140 Pirellula
17	320.2	21.3	110000	1	Continuation (8 of
18	314.8	20.9	110000	1	Continuation (47 o

19	303.8	20.2	2204	15	AF307842	AF307842 Chlamydom
C	301.2	20.0	110000	1	AE017282_15	Continuation (16 o
C	291.8	19.5	110000	1	AY596297_17	Continuation (18 o
22	289	19.2	1518	6	AR480663	AR480663 Sequence
23	287.4	19.1	1515	6	AX571190	AX571190 Sequence
24	287.4	19.1	5910	6	CQ789074	CQ789074 Sequence
25	287.4	19.1	5910	6	AR218933	AR218933 Sequence
26	287.4	19.1	5910	6	BD003845	BD003845 Polynucle
C	287.4	19.1	11298	1	AE007499	AE007499 Streptoco
C	287.4	19.1	34980	6	AX571766	AX571766 Sequence
28	285.8	19.0	3569	1	STRMALMP	J01796 S.pneumonia
C	285.8	19.0	10864	1	AE008556	AE008556 Streptoco
C	284.4	18.9	300349	1	AE017319	AE017319 Desulfovi
32	284.2	18.9	137952	14	SPNEU1909	AL449931 Streptoco
33	265.6	17.7	1377	6	AR586217	AR586217 Sequence
C	211	14.0	110000	1	BA000016_26	Continuation (27 o
C	209.8	14.0	110000	14	CR954200_07	Continuation (8 of
C	209.8	14.0	110000	14	CR954202_06	Continuation (7 of
37	208.2	13.9	1806	15	AY570710	AY570710 Ostreococ
38	200.2	13.3	1731	15	AY081744	AY081744 Arabidops
39	200.2	13.3	1948	15	AY037231	AY037231 Arabidops
C	197.8	13.2	110000	1	CP000024_09	Continuation (10 o
C	196.2	13.1	110000	1	CP000023_09	Continuation (10 o
42	196	13.0	12278	1	AE009809	AE009809 Pyrobacul
C	184.8	12.3	9911	1	AE010052	AE010052 Streptoco
C	184.8	12.3	110000	1	CP000003_10	Continuation (11 o
45	181.4	12.1	2109	15	STDPEPA	X68664 Solanum tub

ALIGNMENTS

RESULT 1	AR393878	Sequence 1	1503 bp	DNA	linear	PAT 18-DEC-2003
LOCUS	AR393878	Sequence 1 from patent US 6617143.				
DEFINITION	AR393878					
ACCESSION	AR393878					
VERSION	AR393878.1	GI:40120997				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 1503)					
AUTHORS	Fukuyama S					
TITLE	Polypeptides having glucanotransferase activity and nucleic acids encoding same					
JOURNAL	Patent: US 6617143-A 1 09-SEP-2003;					
	Novozymes A/S; Bagsvaerd;					
FEATURES	DKX;					
source	Location/Qualifiers					
	1..1503					
ORIGIN	/organism="unknown"					
	/mol_type="genomic DNA"					
Query Match	100.0%;	Score 1503;	DB 6;	Length 1503;		
Best Local Similarity	100.0%;	Pred. No. 3e-243;				
Matches 1503;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	ATGCAACTCCAAACGCGCTTTTGGAAATTTTGCTCCACCCACCACCAAGTTTTCGGGTGCTGG	60			
Db	1	ATGCAACTCCAAACGCGCTTTTGGAAATTTTGCTCCACCCACCACCAAGTTTTCGGGTGCTGG	60			
QY	61	GGGATGGGGCTCTGGGCGCGAGCGCGGTTTGGACTGGTGGCCCATGCGGA	120			
Db	61	GGGATGGGGCTCTGGGCGCGAGCGCGGTTTGGACTGGTGGCCCATGCGGA	120			
QY	121	GCCCGCTGGTGGAGCTTACCGCTGGCCCTACAGTTACGGGACTCGCCGTACCA	180			
Db	121	GCCCGCTGGTGGAGCTTACCGCTGGCCCTACAGTTACGGGACTCGCCGTACCA	180			
QY	181	TCCTTCTCGGCTTTTGGCGGTAAACCGCTATTGGTTGACCCCGAGATGCTATTGAAAA	240			



181 TCCTTCTGGCTTTTGGCGGTAACCGTATTGGTTGACCCGAGATGCTGATTGA AAAA 240  
241 GCCTGGTGGAAACGAGCGGCGCGCGTATCCGACCGCGGTGATTATGCG 300  
241 GCCTGGTGGAAACGAGCGGCGCGCGTATCCGACCGCGGTGATTATGCG 300  
301 TGGCTTTACAGACCGGCTGGCGGCTGTTGGCGCGGCTTTCCGCGGTTTCCGCGCAAG 360  
301 TGGCTTTACAGACCGGCTGGCGGCTGTTGGCGCGGCTTTCCGCGGTTTCCGCGCAAG 360  
361 GCTTCGGCCAGGATGAAGACCGGATGGAAGCCCTTTATCGAGCCGAGCGCTTCGGCTG 420  
361 GCTTCGGCCAGGATGAAGACCGGATGGAAGCCCTTTATCGAGCCGAGCGCTTCGGCTG 420  
421 GAAGACTATGCGCTCTTTATGCGGCTTCAAGACCGGCTTTGAGCGCAAGCCCTGGAACGAG 480  
421 GAAGACTATGCGCTCTTTATGCGGCTTCAAGACCGGCTTTGAGCGCAAGCCCTGGAACGAG 480  
481 TGGAGCCCGAGCTGGCGGACCGGCTGAACCGGCTGCGCCGAGGCGCGCTGAGGAGCTG 540  
481 TGGAGCCCGAGCTGGCGGACCGGCTGAACCGGCTGCGCCGAGGCGCGCTGAGGAGCTG 540  
541 GCGGAGGAGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTATCTGGAATGGGCGCAG 600  
541 GCGGAGGAGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTATCTGGAATGGGCGCAG 600  
601 ACCAGGCTATGCGGATTCGGAATCCAGGGGATTCAGATTCGCGGATGATGCCATCTTTG 660  
601 ACCAGGCTATGCGGATTCGGAATCCAGGGGATTCAGATTCGCGGATGATGCCATCTTTG 660  
661 GCCTTCGATTCCTCAGATGCTGGGCAACCGGCTGCTTACCTCGAGGCGCGATGCG 720  
661 GCCTTCGATTCCTCAGATGCTGGGCAACCGGCTGCTTACCTCGAGGCGCGATGCG 720  
721 AACCCACAGTGGTGGCGGCTTCCGCGGATTCATCTTCGAAACCGGCGAGCTCTGG 780  
721 AACCCACAGTGGTGGCGGCTTCCGCGGATTCATCTTCGAAACCGGCGAGCTCTGG 780  
781 GGCATTCGCTCTATCGCTGGGATGATGATGGAAGGAGCACTTTCGCTGGTGCATTTGCC 840  
781 GGCATTCGCTCTATCGCTGGGATGATGATGGAAGGAGCACTTTCGCTGGTGCATTTGCC 840  
841 CGCATAGGAGTGGCTCAAGCAGTGCACCTGGTGGGATTCGACCACTTCGCGGGTTT 900  
841 CGCATAGGAGTGGCTCAAGCAGTGCACCTGGTGGGATTCGACCACTTCGCGGGTTT 900  
901 GAAGCCTACTGGGAGTTCGCTTTGGCGGCGGCAATGCTGTGAGGGGCGCTGGTCAAA 960  
901 GAAGCCTACTGGGAGTTCGCTTTGGCGGCGGCAATGCTGTGAGGGGCGCTGGTCAAA 960  
961 GCGCCAGGAGAGTGGTTCGCTGGGCGGCGGCAATGCTGTGAGGGGCGCTGGTCAAA 1020  
961 GCGCCAGGAGAGTGGTTCGCTGGGCGGCGGCAATGCTGTGAGGGGCGCTGGTCAAA 1020  
1021 GCGGAGACCTGGGGTGGATCACCCGAGTGGGCTTTCGCGGATGCTTCGGGTTT 1080  
1021 GCGGAGACCTGGGGTGGATCACCCGAGTGGGCTTTCGCGGATGCTTCGGGTTT 1080  
1081 CCGGAGTGAAGATTTTTCGAGTTCGCTTTTTCGCGTGGAGGACCAACGCTTTTTCGCCAC 1140  
1081 CCGGAGTGAAGATTTTTCGAGTTCGCTTTTTCGCGTGGAGGACCAACGCTTTTTCGCCAC 1140  
1141 AACTACCCCGGCAAGGATGCTGGTGTGATGAGCGGACCAACGAGCAACGACCAACC 1200  
1141 AACTACCCCGGCAAGGATGCTGGTGTGATGAGCGGACCAACGAGCAACGACCAACC 1200  
1201 CTGGGATGCTTCGCAAGGCGGCGGAGCGGCTTTCAGTGGGCGCTTACCTGGCC 1260  
1201 CTGGGATGCTTCGCAAGGCGGCGGAGCGGCTTTCAGTGGGCGCTTACCTGGCC 1260  
1261 CGCTATGCGATCCGCTTGTTCGGAATACGAGGTTCGCGGCGCTTTGATCGAGCTGGCC 1320  
1261 CGCTATGCGATCCGCTTGTTCGGAATACGAGGTTCGCGGCGCTTTGATCGAGCTGGCC 1320

1321 TTCAAAAGCCCGCAAGCTGGCTATTGTGCTTTCAGGAGCTGCTGGGCTGGGCCCC 1380  
1321 TTCAAAAGCCCGCAAGCTGGCTATTGTGCTTTCAGGAGCTGCTGGGCTGGGCCCC 1380  
1381 GAGGCGCGCATGAATTTCCCGGAGCGCTGGGAGCAACTGGCGCTGGCGCTACGCCGAA 1440  
1381 GAGGCGCGCATGAATTTCCCGGAGCGCTGGGAGCAACTGGCGCTGGCGCTACGCCGAA 1440  
1441 GCGGACCTTCGAGCGCGGCTTCGCGCGCGGATTCGCGGCGCTTCGCGGAGCCAGCGCGC 1500  
1441 GCGGACCTTCGAGCGCGGCTTCGCGCGCGGATTCGCGGCGCTTCGCGGAGCCAGCGCGC 1500  
1501 GCT 1503  
1501 GCT 1503

RESULT 2  
AY459352  
LOCUS  
DEFINITION  
Thermus aquaticus 4-alpha-glucanotransferase gene, complete cds.  
ACCESSION  
AY459352  
VERSION  
AY459352.1  
GI:38505480  
KEYWORDS  
Thermus aquaticus  
ORGANISM  
Thermus aquaticus  
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
Thermus.  
REFERENCE  
1 (bases 1 to 1503)  
Park, K.-H. and Park, J.-H.  
Direct Submission  
Submitted (05-NOV-2003) Food Science & Technology, Seoul National  
University, San 56-1, Shillim-dong, Kwanak-gu, Seoul 151-742, Korea  
LOCATION/Qualifiers  
1. 1503  
/organism="Thermus aquaticus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:271"  
1. 1503  
/codon\_start=1  
/transl\_table=11  
/product="4-alpha-glucanotransferase"  
/protein\_id="AA223242.1"  
/db\_xref="GI:38505481"  
/translation="MELPRAFLGLLHPTSLPGYGVVGLLEARDLFLKAGGRFW  
QVLPGLPTGYDPSYQAFSAFAGNLYLIDRLAEKGYLVLDKDPGPQGRVDVGMILYA  
WKPALKAAQCGLEKAPKREDFPAFOKEASWLMKQVALFMALKAQGGGLPWNRP  
LPIRRREKAMKEAENALAEVAFAHWQWLFPEAKWALKEAALGIQIIGDPIFV  
AEDSAEVWAHPWFHLDREGRPLVAVGPPDYFSETGQWGNPLVMDLREGRFSPW  
IARLAKALEFHLVRVDHPRGRPEAYWEIPASCTAVEGRWKVASGEKLFDRIOEVFE  
VPIAADLGVITPEVEPLDRYGLPKMVLQFAFHGMENPELPHNYPAHGRVVVTG  
THDNDTLGWRTASPHERAFLEKADWGISFQEEVFPWALMGLCLKLSVARLIATP  
VDVLAIGSEARMNYPGRPSGNWAWRLPQLTOEHAARLLAMAEATGRT"

Query Match 43.1%; Score 647.4; DB 1; Length 1503;  
Best Local Similarity 65.6%; Pred. No. 3.3e-99;  
Matches 980; Conservative 0; Mismatches 506; Indels 9; Gaps 2;  
1 ATGCAACTCAACCGCTTTTGGAAATTTTCTCCACCCACCAAGTTTTCGGGTTCGCTGG 60  
1 ATGCAAGCTTCTAGGGCTTTGGCTCTCTCCATCCACGAGCTTCCCGGGCTTAC 60  
61 GCGATTTGGGCTCTGCGCGGAGGCGGAGCGGTTTTCGCTGCTGGCTGGCGGATGCGGGA 120  
61 GCGTGGGCGCTCTGCGGCTTCGAGGCGCGGAGCTTCTCTCGCTTCTCTAAAGGGCGCGG 120  
121 GCCCGCTGGTGGCAGGTCTTACCGCTGGCGCTTACCAAGTTTACCGGAGCTCGCGTACCAG 180  
121 GCGGCTTCTGGCAGGTCTCTCCCTGGGCGCCACGCGGCTACGCGGAGCTCCCTCTACAG 180  
181 TCCTTCTGGCTTTTGGCGGTAAACCGGTATTGTTGACCCCGAGATGCTGATTGAAAAA 240

```
181 GCCTTCAGCGCTTCGCGGCAACCCCTACTCATAGACTGAGGCCGTTGGCGGAG 240
241 GCGTGGTGAACAACAGCGAAGCGCCCGCGGTATCCGACCCAGCGCGTGAATATGGC 300
241 GCGTACCTGTCTCTAAAGGA-----CCCGCGCTTCCCGGAGGGGCGGTGAGTACGGC 294
301 TGGCTTTACAGACCCCGCTGCGCCCTGTTCGCGGCGGCTTTTCGCGGGTTTGGCGCAAG 360
295 TGGCTCTACCGCTGGAATAGCCCGCCCTTAAGCGCCCTACAGGGTTTCTTGGAAGA 354
361 GCTTGGCGCCAGGATAGACCCCGACTGGAAGCCCTTTATCAGGCGCGGCGCTTCTGCGCTG 420
355 GCTCCAGGAAGAAAGCGGAGGACTTCCCGCGCTTCCAGGAGAAGGAGGCTCTCTGCGCTC 414
421 GAAGACTATCGCTCTTTATGCGCCTCAAGACCGGTTTGACGGCAAGCCCTGGAAGGAG 480
415 AAGGACTACCGCTCTTTCATGCGCCTCAAGCGCCAGCAGCGGGGCTTCCCTGGAACCGG 474
481 TGGAGCCCGAGCTGCGGACCGTGAACCGGCTGCGCCAGGCGCGCTGAGGAGCTG 540
475 TGGCCCTTCCCTGAGCGCGGAGGAGGAGGCCATGAAGGAGCGGAGCGGCCCTG 534
541 GCGAGGAGTGGCCCTTTACGAGTGAATTCAGTGGCTTTTATCTGGAATGGGCGCAG 600
535 GCGGAGGAGTGGCTTCCAGGCTGGAACCGAGTGGCTTCTTTGAGGCTGGAAGGCC 594
601 ACCAAGCCCTATCCGAATCCAAAGGAGTTCAGATATCGCGCATATGCCCATCTTTGTG 660
595 CTCAAGGAGGAGCGGAGGCTTGGGATCCAGATCATCGGGACATGCCCATCTTCGTG 654
661 GCTTGGATCTCAGATGCTGGGCGCAACCGCAGTACTTCTACCTCGAGGCGGATGGC 720
655 GCGGAGGACTCGGCCGAGTCTGGGCGCCACCCCGAGTGGTTCACCTGGAAGGAGGGA 714
721 AACCCACAGTGTGGCGGCGTTCGCGGAGTACTTCTCGAAACCGGCCAGCTCTGG 780
715 AGGCCCTGTGTGTGGGGGTGCGCCCGAGTACTTCTCGAGACCGGCCAGCGCTGG 774
781 GCGAATCCGCTCTATCGCTGGGATGTGAAGAAAGGACAACTTTGCTGTGTCATTTGCC 840
775 GCGAACCCTCTACCGCTGGGAGCTCTGGAGAGAGAGGGTTCCTCTTTTGGATAGCC 834
841 CGCATTAAGGAGTGGCTCAAGCAGTGGCACCTGGTGGCATCAACCATCTTCGCGGGTT 900
835 CGGCTCGCCCAAGGCCCTTGGAGCTATTCACCTGGTGGCGTGACCACTTCGCGGGCTT 894
901 GAAGCCTACTGGGAGTTCGTTTGGCGGCGCCCAATGCTGTGAGGCGGCTGGGTCAAA 960
895 GAGGCTACTGGGAGATCCCGGCTCTGCGCCACCGCGGTGGAGGGCGCTGGGTGAAG 954
961 GCGCCAGGGAGAGCTTTGCTGGGTGGGCGCCCAACTGAGCGATGCGCCCATCAT 1020
955 GCTTCCGGGAGAGCTTTTACCGCATCCAGGAGTTCGCGGAGAGTGCCCATCTCTG 1014
1021 GCGAAGACTTGGGGTGTATCACCCGAGGTGGAGCTTTGCGCATGCTTCGGGTTG 1080
1015 GCGAAGACTTGGGGTGTATCACCCGAGGTGGAGCTTCGCGGAGCGCTACGCGCTT 1074
1081 CCGGCGATGAAGATTTTGCAGTTTGTCTTTT---TCGGGTGAGGACAAAGCTTTTGGCC 1137
1075 CCGGGATGAAGTCTCTGCAATTCGCTTTGACACAGGCGATGGAACCCCTTTCTTCC 1134
1138 CACAATACCCCGCGACGCGAATGTGTGTGTATACAGCGGAACCCACGACACGACCC 1197
1135 CACAATACCCCGCGCATGCGGGGTGTGTGTATACCGGACCCACGACACGACACCC 1194
1198 ACCTGGGATGTGTCCGACCGCGCGGAGGCGGAGGCGCTTCAATGCGGGCTTACCTG 1257
1195 ACCCTGGGCTGTGACGCGAGCGGCTTCCCGCCACGAGCGGCGCTTCTCGGAGCGGTAC 1254
1258 GCGCGTATGGCATCGCTTGTGTGGAATACAGAGTTCGCGGCGCTTTGATCGAGCTG 1317
```

```
1255 GCGGACTGGGGATCAGCTTCCGCGAGGAGGAGGTGCGCTGGGCCCTGATGGGCTT 1314
1318 GCCTTCAAAAGCCCGGCAAGCTGGCTATGTGCTTTGAGGACGCTGCTGGGCTGGGC 1377
1315 TGTTTGAATCGTGGCCAGGCTGGCCATTTACCCGTCGAGACGTTCTGGCCCTGGGC 1374
1378 CCGAGGCGCGCATGAATCTTCCCGGACGCTGGGGGACAACTGGGCGTGGCGCTACGCC 1437
1375 TCGAAGCCCGCATGAATCTACCGGGGCGCCCTTCGGGCAACTGGGCTGGCGGCTCTT 1434
1438 GAAGGCGACTCGAGCCCGCTTGGCGCGGAGACTGCGGCGCTTCGCGGAGGCCA 1492
1435 CCGGGGAGCTCACCCAGGAGCAGCGCGCGCTCTCTGCGCATGGCGGAGGCCA 1489

RESULT 3
E30186
LOCUS Thermophilic amylomaltase. 1503 bp DNA linear PAT 18-JUN-2001
DEFINITION Thermophilic amylomaltase.
ACCESSION E30186
VERSION E30186.1 GI:13025594
KEYWORDS JP 1999046780-A/1.
SOURCE Thermus thermophilus
ORGANISM Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
Thermus.
REFERENCE 1 (bases 1 to 1503)
AUTHORS Yoshinobu, T., Kazutoshi, F., Michiyo, Y., Hiroki, T., Takeshi, T. and
Shigetaka, O.
TITLE Thermophilic amylomaltase
JOURNAL Patent: JP 1999046780-A 1 23-FEB-1999;
EZAKI GLICO CO LTD
COMMENT OS Thermus flavus
PN JP 1999046780-A/1
PD 23-FEB-1999
PF 07-MAY-1998 JP 1998125121
PR
PI YOSHINOBU TERADA, KAZUTOSHI FUJII, MICHIO YANASE, HIROKI TAKADA,
PI TAKESHI TAKADA, SHIGETAKE OKADA
PC C12N15/09, A23G3/00, A23K1/165, A23L1/105, C12N1/21, C12N9/10, PC
C12P19/18//
PC (C12N15/09, C12R1:01), (C12N1/21, C12R1:19), (C12N9/10, C12R1:19),
PC C12N15/00,
PC (C12N15/00, C12R1:01)
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..1503
FT /organism='Thermus flavus'.
FEATURES
source
1..1503
/organism='Thermus thermophilus'
/mol_type='genomic DNA'
/db_xref='taxon:274'

ORIGIN
Query Match 42.5%; Score 638.4; DB 6; Length 1503;
Best Local Similarity 65.0%; Pred. No. 1.1e-97;
Matches 977; Conservative 0; Mismatches 516; Indels 9; Gaps 2;

QY 1 ATGCAACTCCACGCGCTTTTCGAAATTTTGTCTCCACCCACAGTTTTCGGGTGCGTGG 60
DB 1 ATGGAAGTCTCCCGCGCTTTCGGTCTGCTTCTCCACCCACGAGCTCCCGGCCCTTAC 60
QY 61 GGGATGGGGCTCTCGGCGCGGAGCGGAGCGGTTTTCGACTGGCTGGCCGATGCGGA 120
DB 61 GCGCTCGGCTCTCGGCGCGGAGCGGAGCTTCTCGGCTTCTCAAGGAGGCGGG 120
QY 121 GCCCGCTGTGCGAGGTCTTACCGCTGGGCCCTCAAGTTACGGGACTCGCGGTACCA 180
DB 121 GGGCGGTACTGCGAGTCTCTCCCTTGGGCCCGACGGGCTATGGGACTTCCCGCTAC 180
QY 181 TCCTTCTCGGCTTTTGGCGGTAAACCGGTATTTGGTTGACCCCGAGATGCTGATTGAAA 240
```

```

181 TCCTTCAGCGCCTTCGCGGAAACCCCTACCTCATAGACTGAGGCCCTTCGCGGAAAGG 240
241 GGCTGCTGGAACAAACGAGCGCCCGCCGCTATCCGACCCAGCGCGTGGATTATGCG 300
241 GGCCTACGTG-----CGCCTGGAGGACCCCGGCTTCCCCCAAGCGCGGTGACTACGGC 294
301 TGGCTTTACAGACCCGCTGGCCCTGTTGCGCGGGGCTTTCGCGGGGTTTCGGGCAAGG 360
295 CTCCTCTACGCTTGAAGTGGCCCGCTTGAAGGAGGCGCTTTCGGGGCTTCAAGGAAAG 354
361 GCTTCGCGCCAGGATGAAGCCGACTGGAAGCCTTTATCGAGCCGAGCGCTTCTCGCTG 420
355 GCTTCCCGGAGGAGGAGGCGCTTCGCGCGCTTTCGCGGAGGAGGCGCTTGGTGGCTC 414
421 GAAGACTATGCGCTCTTTATGCGCCTCAAGACCCGCTTTGAACGCGCAAGCCCTGGAACGAG 480
415 GAGGACTACGCTCTTTCATGCGCTTGAAGGGGCGCACGGGGGCTTCCCTGGAACCGG 474
481 TGGAGCCCGAGCTGCGCGACCGTGAACCGGCTGCTGCGCCAGGCGCGTGAAGGAGCTG 540
475 TGGCCCTCTTCCCTGCGGAAGCGGAGGAGGAGGCGCTTGAAGGAGGCGGCTTGAAGGAGGCGCTTGG 534
541 GCGGAGGAGTGGCCCTTACGAGTGGATTCAGTGGCTTTTATCTGGAATGGGGCCAG 600
535 GCGGAGGAGTGGCCCTTACGAGCTTTCACGCTTTCACGAGTGGCTTTCGCGCAGTGGGGGGCC 594
601 ACCAAGGCTTATCGCAATCAAGGGGATTCAGATTATCGCGGATATGCGCATTTCTTTGTG 560
595 TTGNAAGCGGAGCGCGGGTGGGCTGCGGATCCGATCATCGGAGCATGCCATCTTCGTG 654
661 GCTTCGATTCCTCAGATGTCGGGCAACCCGAGTACTTCTTACCTCAGAGCGCGATGCG 720
655 GCGGAGGACTCCGCGGAGTCTGGGCGCACCCGAGTGGTTTCACTGACGAGGAGGCG 714
721 AACCCACGCTGCTGGCGGGCTTCGCGGAGTACTTCTCCAAACCGGCGGAGCTCTG 780
715 CGCCCAACGCTGCTGGCGGGTGGCGCGGAGTACTTCTCGGAGACGGCGGCTGG 774
781 GGCATATCCGCTCTATCGCTGGGATGTGATGGAAGGAGCAACATTTGCTGCTGGTCAATGCG 840
775 GGTAAACCCCTTACCGCTGGGAGCTTTGGAGCGGAGGGTCTCTCTCTGATCCGC 834
841 CGCATAAAGGACGCTCAAGCAGTGCACCTGCTGGCGATCAACCACTTCGCGCGGCTT 900
835 CGTCTGAGAGGCGCTCGAGCTTCTCCACCTGCTGGCATAGACCACTTCGCGCGCTT 894
901 GAAGCTACTGGAGGTTCCGTTTGGCGCGCCCAATGCTGTGAGGGCGCTGGGTCAA 960
895 GAGGCTACTGGGAGATCCCGCAAGCTGCCCGCACGCGCGTGGAGGGGCGCTGGGTCAAG 954
961 GCGCCAGGGAGAGAGCTGTTGCTGCGGTGCGGGCCCAACTGAGCGATGCGCCCATCAT 1020
955 GCGCCGGGAGAGAGCTTCTCCAGAGATCCAGAGGCTCTTCGCGAGGTCCCGCTCTC 1014
1021 GCGGAAGACTGGGGGTGATCAACCCCGAGTGGAGGCTTTGCGCGATGCTTGGGGTTC 1080
1015 GCGGAGGACTGGGGGTGATCAACCCCGAGTGGAGGCTTCGCGACGCTTCGCGCTT 1074
1081 CCGCGATGAGATTTTGCAGTTTGTCTTTT---TCGCGTGAAGCAACGCTTTTGGCC 1137
1075 CCGGGATGAAGTCTCTGCAATTTGCGCTTTGACGAGGAGATGGAACCCCTTCTCC 1134
1138 CACAACTACCCCGCGCAGCGCATGTGGTGGTGTACAGCGGAACCCACGACACGACAC 1197
1135 CACAACTACCTCCACGCGCGGTGGTGTCTACACGCGACCCACGACACGACAC 1194
1198 ACCCTGGATGTTTCCGACCGCGCGAGGCGCGAGCGGCGCTTCAATGGGCGCTACCTG 1257
1195 ACCCTGGCTGTGACCGCAGCGCCACCCCGCGAGAGGCGCTTCAATGGCGCGTACCTG 1254
1258 GCGCGTATGGCATCCGTTGTTGTCGGAATACAGAGTGGCGGGCGCTTTCATCGAGCTG 1317

```

```

1255 GCGGACTGGGGATCACCTTCGCGAGGAGGAGGAGTGGCCCTGGGCCCTGATGACCTG 1314
1318 GCCTTCAAAGCCCGCGCAAGCTGCTATTGTCCTTTTCAGGACGCTGCTGGGCTGGGC 1377
1315 GGGATGAATGCTGCTGGCGCGCTCGCGCTTACCCGCTGTCAGGACGCTCTGGCCCTGGGC 1374
1378 CCGAGGCGCGCATGAATCTCCCGGACGCTGGGGGACAACTGGGCGTGGGCGCTACGCG 1437
1375 ACGAGGCGCGGATGAATACCCCGGAGGCGCTCGGGGACTGGGCTTGGGCTCTCTC 1434
1438 GAAGCGGACTCGAGCCCGGCTGCGCGGAGCTGCGGCGCTGCGCGGAGCGACCGAG 1497
1435 CCGGGGAGCTTCCCGGAGCACGCGGCGAGGCTTAGGGCCATGCGCGGAGCGACGAA 1494
1498 CG 1499
1495 CG 1496

```

## RESULT 4

AB016244  
LOCUS  
DEFINITION  
Thermus aquaticus gene for amylomaltase, complete cds.  
ACCESSION  
AB016244.1 GI:3721879  
VERSION  
amylomaltase.  
KEYWORDS  
Thermus aquaticus  
SOURCE  
Thermus aquaticus  
ORGANISM  
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
Thermus.

2085 bp DNA linear BCT 24-MAR-1999  
Thermus aquaticus gene for amylomaltase, complete cds.

AB016244.1 GI:3721879

amylomaltase.

Thermus aquaticus

Thermus aquaticus

Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;

Thermus.

1 (sites)

Terada,Y., Fujii,K., Takaha,T. and Okada,S.

Thermus aquaticus ATCC 33923 amylomaltase gene cloning and

expression and enzyme characterization: production of cycloamylose

Appl. Environ. Microbiol. 65 (3), 910-915 (1999)

10049841

2 (bases 1 to 2085)

Terada,Y., Fujii,K., Takaha,T. and Okada,S.

Direct Submission

Submitted (14-JUL-1998) Yoshinobu Terada, Ezaki Glico Co., Ltd.,

Biochemical Research Laboratory, Utsunomiya 4-6-5, Nishiyodogawa-ku,

Osaka, Osaka 555-8502, Japan [E-mail:terada-yoshinobu@lico.co.jp,

Tel:81-6-477-8425, Fax:81-6-477-8271]

Location/Qualifiers

1..2085

/organism="Thermus aquaticus"

/mol\_type="genomic DNA"

/strain="ATCC33923"

/db\_xref="taxon:271"

514..2016

/gene="malQ"

514..2016

/gene="malQ"

/codon\_start=1

/transl\_table=11

/product="amylomaltase"

/protein\_id="BA03728.1"

/db\_xref="GI:3721880"

/translation="MELPRAFLGLLHPTSLPGYGVGLREARDFLFLKEAGGRYV

QVLPALPGYGDSPYQSFSAFAGNPLYDLRLAERGVYRLDPGPGPQGRVDYGLIYA

WKWPLKEAFKGFKEKASPEREAPFAPREAWLEDYALFWALKGAHGLPWNRP

LPLKREKALREKASALAEVAFHATPTOMLFRONGALKAEEALGIRIIGDMPLEFV

ARDSAEVVAHPEWFLHDEGRPTVAVGVPDFTSTGQRGNPLVRWDVLERGFSFW

IRRLALELHFLVRIDHFRGEAYWEIPASCTAVEGRWVKAPGKLFQKIQEVPGE

VPVLAEDLVITPEVALDRFLGPMKVLQFADDDGEMENPLPHNYPHAGRVVYTG

THDNDTLGWRTATPEKAPMARYLDWGITFRSEEVPMALHLMGKMSVARLAVYP

VQDVLAGSEAMNYPGRPSGNWRLPLGELSPEHGRARLRAVAETRL"

## ORIGIN

Query Match 42.5%; Score 638.4; DB 1; Length 2085;  
Best Local Similarity 65.0%; Pred. No. 1e-97;  
Matches 977; Conservative 0; Mismatches 516; Indels 9; Gaps 2;



/locus\_tag="TTC0847"  
 complement (290. .1864)  
 /locus\_tag="TTC0847"  
 /EC\_number="4.1.3.12"  
 /codon\_start=1  
 /transl\_table=11  
 /product="2-isopropylmalate synthase"  
 /protein\_id="AAS81191.1"  
 /db\_xref="GI:46196776"  
 /translation="MVEILDTLRDGTGGVSLSDVDKVAIARRLAAFGHILIEGGW  
 PGNXPDMRFGVGLGAALAFGATRRKGLAEEEDPSVLALLEATPPVVVLPFG  
 KSWLHVLEALSLSENLMIEETVAFFARRGRVYDAEHFPGDKYKEDPAVALATL  
 ENAFRGADTLVLCDDGLPEVIAITKAVVERFPGVGRVGHPHNDADLANALAL  
 AVRAGATHVQGTINGERCNNLNSFLTLVFKYGIPIAISPEKRELEVSFHDGR  
 RANLTNNRPAAYFAAGVAGVHVAFLNPRTEYHIPPWVQNRRLFLVSDVGR  
 SNLLAKLOELGVLDSEAKRRLLEKALEVEGYAPGEASFYLLAHLRLKGGSLPFS  
 VEGSFVYVGRGLSTAWAETATRVVVGESLQHTAAESPFGVSALDRAFKAVLOFYP  
 EUSEVELTKVKRILSGQAGTNSGVVRVEMKRGGERPATVGNASENILEASLKALTD  
 GYAYALLYPVATPRGA"  
 complement (1858. .2385)  
 /locus\_tag="TTC0848"  
 complement (1858. .2385)  
 /locus\_tag="TTC0848"  
 /EC\_number="2.3.1.1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="acetyltransferase"  
 /protein\_id="AAS81192.1"  
 /db\_xref="GI:46196777"  
 /translation="MVRIRKAGLEDLPGVARVLVDWTRATRVGVVPFAFLEGISYEQ  
 AERWAQRLKTPWFLFAESSEGVVGFATPGDPGRGFYEHLLGGVLLGEREIELGGAK  
 RKLGLALFHEGALLQAEYGRMLVWLKENPKRGFYEHLLGGVLLGEREIELGGAK  
 LWEVAYGFDLGGHKW"  
 complement (2411. .3973)  
 /locus\_tag="TTC0849"  
 complement (2411. .3973)  
 /locus\_tag="TTC0849"  
 /EC\_number="4.1.3.12"  
 /codon\_start=1  
 /transl\_table=11  
 /product="2-isopropylmalate synthase"  
 /protein\_id="AAS81193.1"  
 /db\_xref="GI:46196778"  
 /translation="MEKERHIRFDTLRDGEQSPGVALSIDQKLEIAQALQNLVDI  
 IEAGFVSGMEFEAEVRIATEVKGPITIAALARTHTLIDIOAKAKLAEKPRTHVPT  
 SASKVHLMYKTEEEVLEMAWRVYARYVDDVEFSAQDVWRADWEFKRLIYEVA  
 IEAGATINIPDTGYTGALIRIRIDEVVRGRDVIISTHDDLGWATANALA  
 RVNGAGQIETTINGERAGNCALAEVVMALYVRDWMYKATRINTREIRVRSRLVE  
 GTVGMPPVNNKATVGNAPAHESGIGHQGVIKHATYIEMDAELIGRRPAVLVGRHS  
 GRAAFKALDELIGVLDSEBEVKLFAFKFIEAKKGPLSAEELQALVESEREPTSHF  
 FQLEHVQFSGSGLLPATVYKVTPOGERVATHTGDPDAVYKAIQOALIGRLPELEL  
 YRVEAITGSTEARLQGVTVRLURLGELQAVGVGVSPDIIEASALAFDAAKGLASGRAT  
 HPPSIEVHRGV"  
 complement (4031. .5044)  
 /locus\_tag="TTC0850"  
 complement (4031. .5044)  
 /locus\_tag="TTC0850"  
 /EC\_number="1.1.1.86"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ketol-acid reductoisomerase"  
 /protein\_id="AAS81194.1"  
 /db\_xref="GI:46196779"  
 /translation="MKIYYEHDADLGFILGKVAVLGFGSQGHAAHLNLKDSGVDRV  
 GLKGRSWEKAAAGLRVLVFAAREADVVMVLLPDEKQAVGVPAALVEHPNLKEGGA  
 LFAHGFNVHFGQIKPRKDLIDVMVAFKFGHVLRYSEYQKSGVPALVAVHQDASGA  
 FPTALYAKAIGARAGVATTPDETFDLFGQVALCGGLTRLIRAGFETLVEAGY  
 PPMWAFVETHEVKLIYDLIYEAGLKMRYISINISYAGDYTRDGLAVPLEETKRRMR  
 EILRQIQSGFAEWMLENQVGSVLEAKRWAHAHIEEVGSLRLAMPPKLKARVME  
 EVG"  
 complement (5041. .5553)  
 /locus\_tag="TTC0851"  
 complement (5041. .5553)

CDS	/locus_tag="TTC0856" 9613..10314 /locus_tag="TTC0856" /codon_start=1 /transl_table=11 /product="prepilin-like protein" /protein_id="AAS81200.1" /db_xref="GI:46196785" ORF1:10314-11989 /gene="comZ"	54315	GGCAACCCCTTTACCGCTGGGACGTTTGGAGCGGAGGGGTTCTCTCTTGGATCCGC	54255
Qy	841	CGCATAGGCGAGTCGCTCAAGCAGTGCACCTGGTGGGATCGACCACTTCCGGCGGGTTT	900	
Db	54255	CGTCTGGAGAGGCCCTTGGAGCTCTTCACTGGTGGGATAGACCACTTCCGGCGCTTT	54196	
Qy	901	GAAGCCTACTGGGAGGTTCCGTTTGGCGGCCCAATGCTGTGGAGGGGCGCTGGGTCAAA	960	
Db	54195	GAGGCTTACTGGGAGATCCCGCAAGCTGCCACGGCGGTGGAGGGGCGCTGGGTCAAG	54136	
Qy	961	GCCTCAGGGGAGAGCTGTTTGTCTGGGTGGGCGCCCACTGAGCGATGCGCCCATCAT	1020	
Db	54135	GCCCCGGGGAGAGCTCTTCCAGAGATCCAGAGGCTTTCGGCGAGGTCCTCCGCTCCTC	54076	
Qy	1021	GCCGAGACCTGGGGGTGATCACCCCGAGGTGGAGGCTTTGGCGGATGGCTTTCGGGGTTC	1080	
Db	54075	GCCGAGGACCTGGGGGTGATCACCCCGAGGTGGAGGCTTTCGGCGGACCGCTTTCGGGCTT	54016	
Qy	1081	CCCGCATGAAGATTTTGCAGTTTGTCTTTT---TCCGGTGAAGGCAACGCTTTTTCGCC	1137	
Db	54015	CCCGGATGAAGTCTCTGAGTTTGCCTTTTGAACGAGGATGGAACCCCTTCTCTCCTC	53956	
Qy	1138	CACAACTACCCCGCGCACCGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1197	
Db	53955	CACAACTACCCCGCGCACCGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	53896	
Qy	1198	ACCTGGGATGGTTCGCGACCGCGCGAGGCGCGAGGCGCTTTCATGCGGGCTACCTG	1257	
Db	53895	ACCTGGGCTGGTACCGCACCGGCCACCCCCACGAGAGGCTTTCATGCGGGCTACCTG	53836	
Qy	1258	GCCCGCTATGSCATCCGTTTGTTCGGAATACAGAGTCGCGGGCGCTTTTGTATCGAGCTG	1317	
Db	53835	GCGGACTGGGGGATACACCTTCGCGGAAGAGAGGAGGTGCCCTGGGCGCTGATGCACCTG	53776	
Qy	1318	GCCTTCAAAAGCCCGGCCAAGCTGTGTATTTGTGCTTTTCAGAGAGCTGCTGGGCTGGG	1377	
Db	53775	GGGATGAAGTCCGTGGGCGCGGCTCGCCGTTTACCCGCTGAGGAGCTCTTGGGCGCTGGG	53716	
Qy	1378	CCCGAGGCGCGCATGAACCTTCCCGGACCGCTGGGGGAGCAACTGGGGGTGGCGCTAGCGC	1437	
Db	53715	AGCGAGGCGCGGATGAACCTACCGGGAAGGCCCTTCGGGGAACCTGGGCTGGGCTCCTC	53656	
Qy	1438	GAAGCGCACTTCGAGCCCGCTTGTGGCCGCGGAGCTGCGGGCCCTTGGCCGAGGCCAG	1497	
Db	53655	CCGGGGAGCTTTCCTCCGAGACGCGGGCGAGGCTTAGGGCCATGGCCGAGGCCACGAA	53596	
Qy	1498	CG 1499		
Db	53595	CG 53594		
RESULT 6	BD179859	1500 bp	DNA linear	PAT 15-MAY-2003
LOCUS	BD179859	Highly thermophilic bacterium-derived protein and gene encoding it.		
DEFINITION	BD179859	Highly thermophilic bacterium-derived protein and gene encoding it.		
ACCESSION	BD179859	Highly thermophilic bacterium-derived protein and gene encoding it.		
VERSION	BD179859.1	GI:30790777		
KEYWORDS	JP 2002325574-A/350.			
SOURCE	Thermus thermophilus			
ORGANISM	Thermus thermophilus			
REFERENCE	1 (bases 1 to 1500)			
AUTHORS	Kuramitsu, N. and Yokoyama, S.			
TITLE	Highly thermophilic bacterium-derived protein and gene encoding it			
JOURNAL	Patent: JP 2002325574-A 350 12-NOV-2002;			
COMMENT	THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH			
	OS Thermus thermophilus			
	PN JP 2002325574-A/350			
	PD 12-NOV-2002			
	PP 23-FEB-2001			
	PI NARUKI KURAMITSU, SHIGEKI YOKOYAMA			

Query Match	42.5%; Score 638.4; DB 1; Length 273285;		
Best Local Similarity	65.0%; Pred. No. 5e-98;		
Matches 977; Conservative 0; Mismatches 516; Indels 9; Gaps 2;			
Qy	1	ATGCAACTCCAAACCGCTTTTGGAAATTTTGTCTCCACCACCAAGTTTTCGGGTCTGCTGG	60
Db	55089	ATGGAGCTTCCCGCGCTTTCGGTCTGCTTCTCCACCACAGGCTTCCCGGCCCTTAC	55030
Qy	61	GGGATTTGGGCTTGTGGCGCGAGGCGAGCGGTTTTTGGACTTGGCTGGGCCGATCGGGA	120
Db	55029	GGCGTGGCGTCTTGGCGCAGGAGGCGCGGAGCTTCTCGGCTTCTTCAAGGAGGCGGG	54970
Qy	121	GCCCGCTGGTGGCAGGCTTACCGCTGGGCGCTTACAGTTTACGGGCACTCGCGTACCA	180
Db	54969	GGGCGGTACTGGCAGGCTTCTCCCTTGGGCGCCACGGGCTATGGCAGCTCCCTTACC	54910
Qy	181	TCCTTCTCGGCTTTTGGCGGTAAACCCGTAATTTGGTTGACCCGAGATGCTGATGA	240
Db	54909	TCCTTACGCGCTTCGCGGAACCCCTACTCATAGACCTGAGGCCCTTCGCGGAAGG	54850
Qy	241	GGCTGGCTGGAACAAAGCGAAGCGCCCCCGCGTATCCGACCCAGCGGCTGATATG	300
Db	54849	GGCTACGTG-----CGCCTGGAGGACCCCGGCTTCCCCAAGCGCGGTGGACTAC	54796
Qy	301	TGGCTTTACAGACCCGCTGGCGCTTGTGGGCGGCTTTCGCGGGTTTCGGGCAAG	360
Db	54795	CTCCTCTACGCTGGAAGTGGCGCGCCCTGAAGAGGCGCTTCCGGGCTTCAAGGA	54736
Qy	361	GCTTTCGGCCAGGATAGACCGCACTGGAAGCCTTTATCGAGCCGAGGCTTCTGGCTG	420
Db	54735	GCCTCCCGGAGGAGCGGAGGCTTTCGCGCTTTCGGGAGAGGAGGCTTGGTGGCTC	54676
Qy	421	GAAGACTATCGCTTTTATGGCCCTCAAGACCCCGTTTGAACGCAAGCCCTGGAAC	480
Db	54675	GAGGACTACGCCCTTTCATGGCCCTGAAGGGGCGCACGGGGGCTTCCCTGGAACCG	54616
Qy	481	TGAGCCCCGAGCTGGCGACCGTGAACCGGCTGCCCTGGCCAGGGCGCTGAGGAGCTG	540
Db	54615	TGGCCCCCTTCCCTGGGGAAGCGGGAAGAGAGGCCCTTAGGGAGCGAAGACGCTTG	54556
Qy	541	GCGGAGAGGTGGCCCTTACGAGTGGATTCAGTGGCTTTTATCTGAAATGGGCGCAG	600
Db	54555	GCGGAGAGGTGGCTTCCAGGCTTCAACCATGCTTCTTTCGCGCAGTGGGGGCC	54496
Qy	601	ACCAAGGCTTATGCCGAATCAAGGGGATTCAGATTATCGCGATATGCCATCTTTGTG	660
Db	54495	TTGAAGCGGAGCGCGGGCTTGGGCATCCGATCATCGGGACATGCCATCTTCGTG	54436
Qy	661	GCCTTCGATTCCTCAGATGCTGGGCAACCCGCACTTCTTCACTCGAGGCGGATGGC	720
Db	54435	GCGGAGGACTCCCGCGAGGCTTGGGCGCAACCCGAGTGGTTTCACTCGACGAGGAGG	54376
Qy	721	AACCCACAGTGTGGCGGCGTTTCGCGGGACTTCTTTCGGAACCGGCCAGCTCTGG	780
Db	54375	CGCCCCACGCTGTGGCGGGGTGGCCCCGACTTCTTCTCGGAGACGGGCGGCTG	54316
Qy	781	GGCAATCCGCTCTATCGCTGGGATGTGATGGAAGGGGCAACTTTTCCCTGGTGCATTC	840



PC C12N15/09, C12N15/09, C07K14/195, C12N1/15, C12N1/19, C12N1/21, PC  
C12N5/10,  
PC C12N9/88, C12P21/02, C12N9/88, C12R1/01, (C12N15/09, C12R1/01),  
PC (C12P21/02, C12P21/01, C12N15/00, C12N15/00, C12N5/00, (C12N15/00,  
C12R1/01)  
CC Highly thermophilic bacterium-derived protein and gene CC  
encoding it  
FH Key Location/Qualifiers  
FT CDS (1)..(1500).  
FEATURES  
source Location/Qualifiers  
1..1500  
/organism="Thermus thermophilus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:274"  
ORIGIN  
Query Match 42.4%; Score 636.8; DB 6; Length 1500;  
Best Local Similarity 65.0%; Pred. No. 2e-97;  
Matches 976; Conservative 0; Mismatches 517; Indels 9; Gaps 2;  
QY 1 ATGCAACTCCAAAGCGCTTTTGGAAATTTGTCTCCACCCACCAAGTTTTCGGGTTCGCTGG 60  
DB 1 ATGAGCTTCCCGCGCTTTTCGGTCTCTTCTCCACCCACGAGCTTCCCGGCCCTAC 60  
QY 61 GGATTTGGGCTTTGGGCGCGAGGCGAGCGGTTTGGACTGGCTGGCGGATGGGGA 120  
DB 61 GGCGTGGGCTTCTGGGCGAGGAGCGCGGACTTCTCTCGCTTCTCAAGGAGGCGAGG 120  
QY 121 GCGGCTGGTGGAGGCTTACCGCTGGCGCTTACCAAGTTTACGGCGACTCGCGTACCAG 180  
DB 121 GGGCGGTAATGGGAGGCTTCTCCCTTGGGCGGCGGCGGCTATGGCGACTTCCCTACAG 180  
QY 181 TCCTTCTCGGCTTTGGCGGTAAACCGTATTTGGTGTGACCCCGAGATGCTGATGAAAAA 240  
DB 181 TCCTTACGCTTTCGCGGAAACCCCTACTCATAGCTGAGGCCCTTCGCGGAAAGG 240  
QY 241 GGCTGGCTGGAACAAAGCGAAGCGCCCGCGCTATCCGACCCAGCGGCTGGAATATGGC 300  
DB 241 GGCTACGCTG-----CGCTGGAGGACCGCGCTTCCCGGCGGCGGCTGGAATACGCG 294  
QY 301 TGGCTTTACAGACCGCTGGCGGCTTGGCGGCGCTTTCGGGCGGCTTTCGGGCGAGG 360  
DB 295 CTCTCTACGCTGGAAGTGGCGCGCTTGAAGGAGGCTTTCGGGCGGCTTCAAGGAAAG 354  
QY 361 GCTTCGCGCCAGGATAAGACCGACTGGAAGCTTTATCGAGGCGGAGCGCTTCTGGCTG 420  
DB 355 GCCTCCCGGAGGAGCGGAGGCTTTCGCGCTTTCGGGAGAGGAGGCTTGGTGGCTC 414  
QY 421 GAAGACTATGCGCTTTTATGCGCTTCAAGACCGGTTTGAACCGGCTTTCGGAACGAG 480  
DB 415 GAGGACTACGCGCTTCTCATGGCGCTGAAGGGGCGCACGGGGGCTTTCCTGGAACCGG 474  
QY 481 TGGAGCCCGAGCTGGCGGACCGTGAACCGGCTGCGCTGGCGGCGGCGCTGAGGAGCTG 540  
DB 475 TGGCGCTTCCCTTCGCGAAGCGGGAAGAGAGGCTTTCGGGAGGCGGAAAGCGCTTG 534  
QY 541 GCGGAGAGGTGGCGCTTTACAGTGGATTTCAGTGGCTTTTATCTGGAATGGGCGCAG 600  
DB 535 GCGGAGAGGTGGCTTCCAGCGCTTACCGAGTGGCTTCTTCGCGAGTGGGGGCGC 594  
QY 601 ACCAAGGCTATGCGGAATCAAGGGGATTCAGATTATTCGCGGATATGCCCATCTTTGTG 660  
DB 595 TTGAAGCGGAGGCGGAGGCGTTGGGATCCGGATCATCGGGGACATGCCCATCTTCGTG 654  
QY 661 GCCTTCGATTCAGATGCTGGGCGAACCAGCGAGTACTTCTACCTCGAGCGGCGATGGC 720  
DB 655 GCGGAGGACTCCGCGAGGCTTGGGCGGCGGCTTTCAGCTTTCAGCTGAGAGAGGCG 714  
QY 721 AACCCGAGGTGGGCGGCTTTCGCGGAGTACTTCTCGGAAACCGGCGAGCTTGG 780  
DB 715 GCGGCGAGGTGGGCGGCGGCTTTCGCGGAGTACTTCTTCGAGAGCGGCGAGGCTGG 774  
QY 781 GGCAATCCGCTCTATCGCTGGGATGTGATGGAAGGGAACAATTTTGGCTGTGATTCGCC 840

DB 775 GGCAACCCCTTTACCGCTGGGACGTTTGGAGGGGGTTCCTCTCGATCCGC 834  
QY 841 CGCATAGGACATGCTCAAGCAGTGCACCTGGTGGGATGACACCACTTCGCGGGTTT 900  
DB 835 CGTCTGGAGAAAGCGCTTTCACCTGGTGGCATAGACCACTTCGCGGGTTT 894  
QY 901 GAAGCTACTGGGAGGTTCCGTTTGGCCCGGCCCAATGCTGTGGAGGGGCTGGGTCAAA 960  
DB 895 GAGGCTACTGGGAGATCCCGCAAGCTGCCCGCAGCGGTGGAGGGGCGCTGGGTCAAG 954  
QY 961 GCGCCAGGGGAGAAAGCTGTTTGTGGGTGCGGCGCCCAACTGAGCGATGCGCCCATCAT 1020  
DB 955 GCGCCGGGGGAGAAAGCTTTCAGAGATCCAGAGGCTTTCGCGCGAGGTCCCGCTCTC 1014  
QY 1021 GCGGAAGACTGGGGGTGATCAACCCCGAGGTGGAGGCTTTGGCGCATGGCTTCGGGTTT 1080  
DB 1015 GCGGAGGACTGGGGGTGATCAACCCCGAGGTGGAGGCGCTTCGCGCAACGCTTCGGGCTT 1074  
QY 1081 CCGGATGAAGATTTTGCAGTTTGTCTTTT---TCGGGTGAGGACCAACGCTTTTGGCCC 1137  
DB 1075 CCGGATGAAGGTCCTGAGTTTCGCTTTGACGACGGATGGAAGAAACCCCTTCCTCCCC 1134  
QY 1138 CAACAATCAACCGCGCAACGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1197  
DB 1135 CAACAATCAACCGCGCAACGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1194  
QY 1198 ACCCTGGGATGTTTCGCGCACCGCGCGCGAGGCGCGGCGCTTTCATGCGGGCTTACCTG 1257  
DB 1195 ACCCTGGGCTGTGTACGCGCACCGCGCACCGCGCGCTTTCATGCGGCGGTACCTG 1254  
QY 1258 GCGCGTATGGCATCCGTTGTTGTGCGAATAGAGTGCAGGCGCGCTTGTGTGTGTGTGTGT 1317  
DB 1255 GCGGATGGGGATCACTTTCGCGGAGAGGAGGAGTGCCTTGGGCGCTGTGTGTGTGTGT 1314  
QY 1318 GCCTTCAAAAGCCCGCAAGCTGTGTATTTGCTTTCAGGACGCTGTGTGTGTGTGTGTGT 1377  
DB 1315 GGGATGAGTCCGTGGCGCGGCTTCGCGCTTACCGGTGCGAGACGCTTCTGGCGCTTGGC 1374  
QY 1378 CCGAGGCGCGCATGAATTCCTCCGCGAGCGCTGGGGAACAATGGGGGTGGCGTACGCC 1437  
DB 1375 AGCGAGGCGCGGATGAATACCGCGGAAAGGCGCTTCGCGGAACTGGGCGCTGGCGCTC 1434  
QY 1438 GAAGGCGACTTCAGCGCGCTTGGCGCGGAGTGCAGGCGCTTGGCGGAGGCGGCGAG 1497  
DB 1435 CCGGGGAGGCTTTCGCGGAGACCGGCGCGAGGCTTAGGGCCATGGCGGAGGCGGAA 1494  
QY 1498 CG 1499  
DB 1495 CG 1496

RESULT 7  
AX801458  
LOCUS AX801458 1503 bp DNA linear PAT 13-OCT-2003  
DEFINITION Sequence 253 from Patent WO03002728.  
ACCESSION AX801458  
VERSION AX801458.1 GI:37654049  
KEYWORDS  
SOURCE Thermus thermophilus  
ORGANISM Thermus thermophilus  
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
Thermus.  
REFERENCE  
1  
AUTHORS van der Maarel, M.J., Dijkhuizen, L., Binnema, D.J., van der Veen, B.A.  
and Vos, A.  
TITLE Prevention of retrogradation of starch  
JOURNAL Patent: WO 03002728-A 253 09-JAN-2003;  
Toegepast-Natuurwetenschappelijk Onderzoek TNO (NL)  
FEATURES  
Location/Qualifiers  
1..1503  
/organism="Thermus thermophilus"  
/mol\_type="unassigned DNA"





**Tue Jan 17 09:28:39 2006**

ug-10-618-976-1.rge

AP008226_14	1400001	1510000
AP008226_15	1500001	1610000
AP008226_16	1600001	1710000
AP008226_17	1700001	1810000
AP008226_18	1800001	1849742

Continuation (12 of 19) of AP008226 from base 1100001 (AP008226 Thermus thermophilus)

Query Match	42.4%	Score 636.8	DB 1	Length 110000
Best Local Similarity	65.0%	Prod. No. 1.1e-97		
Matches	976	Conservative	0	Mismatches 517; Indels 9; Gaps 2
QY	1	ATGCAACTCCAAACGGCTTTTGGAAATTTTGTCTCACCCACACAGTTTTCGGGTTCGCTGG	60	
DB	103076	ATGGAGCTTCCCGCGCTTTTCGGTCTGCTTCTCCACCCACAGAGCTCCCGGCCCTTAC	103017	
QY	61	GGGATTGGGGCTCTGGCGCGAGGCGAGCGGTTTTGGACTGGCTGCGCGATCGCGGA	120	
DB	103016	GGCGTGGCGTCTTGGGACAGAGCGCCGGACTTCTCTCGCTTCTCAAGAGGACAGG	102957	
QY	121	GCOCGTGGTGGAGGCTTTACCGCTGGGCCCTACCACTAGCGGAGCTTCGCCGTACCAG	180	
DB	102956	GGCGGTACTGGCAGGTCTCCCTTGGGCCCAAGGCTATGGGACTTCCCGCTTACCAG	102897	
QY	181	TCCTTCTCGGCTTTTGGCGGTAAACCGTATTTGGTTGACCCGAGATGCTGATTGAAAA	240	
DB	102896	TCCTTACGCGCTTTCGCGGAAACCCCTACCTCATAGACTTGAGGCCCTCGCGAAGG	102837	
QY	241	GGCTGGCTGGAAACAAAGCGAAGCGCCCGCGCTATCCGACCCAGCGGTGATTTATGCG	300	
DB	102836	GGCTACGTG-----CGCTGGAGACCCCGGCTTCCCGGAGGCGGTGACTACGCG	102783	
QY	301	TGGCTTTACCAAGACCGCTGCCTCTGTGTCGGCGGCTTTTCGGGGTTTCGGGCAAG	360	
DB	102782	CTCCTTACGCTGGAAGTGCCCGCTTGAAGAGGCGCTTCGGGGCTTCAAGGAAAG	102723	
QY	361	GCTTCGGCCAGGATAAGACCGACTGGAAGCTTTATCGAGCGCGAGCGCTTCCTGGGTG	420	
DB	102722	GCCTCCCGGAGGACGCGGAGGCGCTTTCGCGGCTTCCGGGAGAGGAGGCTTGTGGCTC	102663	
QY	421	GAAGACTATGCGCTCTTTATGGCCCTCAAGACCCCGTTTGAAGCGAAGCCCTTGAACGAG	480	
DB	102662	GAGGACTACGCCCTCTTATGAGCCCTTGAAGGGGCGCAGCGGGGCTTCCCTGGACCGG	102603	
QY	481	TGGAGCCCGAGCTGCGGACCGTGAACCGGCTGCCCTGGCGGCGCCGTGAGGAGCTG	540	
DB	102602	TGGCCCTTCCCTTCGGAAGCGGGAAGAGAGGCGCTTAGGAGGCGGAAAGCGCCTTG	102543	
QY	541	GCCGAGGAGTGGCCCTTTACAGTGGATTCACTGGCTTTTATCTGGAATGGGGCCAG	600	
DB	102542	GCCGAGGAGTGGCCCTTCCACGCTTACCCAGTGGCTCTTCTTCGCGAGTGGGGGCGC	102483	
QY	601	ACCAAGGCTATGCGGAATCCAAAGGGATTCAGATTATCGCGGATATGCCCATCTTTGTG	660	
DB	102482	TTGAAGCGGAGGCCGAGCGGTTTGGGATCCGGATCATCGGAGCATGCCATCTTCGTG	102423	
QY	661	GCCTTCGATTCTCAGATGTCTGGGCAACCCGCGATCTTCTACCTCGAGGCGGATGCG	720	
DB	102422	GCCGAGGACTCCGCGAGTCTGGGCCACCCCGAGTGGTTTTCACCTGAGAGGAGGCGG	102363	
QY	721	AACCCCAAGTGGTGGCGGCGTTTCGGCGGACTTCTTCGGAACCGGCGAGCTCTGG	780	
DB	102362	GCCCCAAGTGGTGGCGGCGTGGCCCCCGGCTACTTCTCGAGAGCGGCGAGCGCTGG	102303	
QY	781	GGCAATCCGCTCTATCGCTGGGATGTGATGGAAAGGACAACTTTCCTCTGGTGCATTTGCC	840	
DB	102302	GGCAACCCCTTTACCGCTGGGACGTTTGGAGCGGAGGGGTTCTCTTCTGATCCCG	102243	
QY	841	GCATAAGGCACTCGCTCAAGCAGTGTCCACCTGGTGGCATGACCACTTCGCGGGTTT	900	
DB	102242	CGTCTGGAGAAAGCCCTTGAGGCTTCTCCACCTGGTGGCATAGACCACTTCGCGGCTT	102183	
QY	901	GAAGCCTACTGGAGGTTCCGTTTGGCGGCGCCCAATGCTGTGAGGGGCGCTGGGTCAA	960	

Db	102182	GAGCCCTACTGGGAGATCCCGCAGCTGCCCCACGCGGGTGGAGGGCGCTGGGTCAAG	102122
Qy	961	GCCCCAGGGGAGAGCTGTTTGCTGCGGTGCGGCCCAACTGAGCGATGCGGCCCATATT	1020
Db	102122	GCCCGGGGAGAAAGCTCTTCCAGAAATCCAGAGGTCTTTCGGCGAGGTCCCCGTCCTC	102063
Qy	1021	GCCGAAGACTGGGGGTGATACACCCCGAGGTGAGGCTTTGCGCATGGCTTCGGGTTC	1080
Db	102062	GCCGAGACCTGGGGGTGCATCACCCCGAGGTGAGGCCCTTCGGGACCGCTTCGGCCTT	102003
Qy	1081	CCCGGCATGAAGATTTTGCAGTTTGTCTTTT---TCCGGTGAAGGACAACGCCCTTTTGGCC	1137
Db	102002	CCCGGATGAAGTCTCTGCAGTTCGCCCTTTGACGACGGGATGGAACCCCTTCTCCGCC	101943
Qy	1138	CACAACTACCCCGCGCAGCGCAATGTGTGTGTATCAGCGGAACCCACGACNACACACACC	1197
Db	101942	CACAACTACCTGCCCCACGGCGGGTGTGTCTACACCGGCACCCACGACNACACACACC	101883
Qy	1198	ACCTGGGATGTTCCGCACCGCGCGAGGCCCTTCATCGGGCCCTACCTG	1257
Db	101882	ACCTGGGCTGTACCGCACCGCCACCCCCACGAGAGGCTTCATGGCGGTACCTG	101823
Qy	1258	GCCCGCTATGCGATCCGTTGTTGTTCGGAATACGAGGTGCGGGCGCTTTGATCGAGCTG	1317
Db	101822	GCGNCTGGGGATCACCTTCGGGAGAGGAGAGGTGCCCTGGGCCCTGATGACACCTG	101763
Qy	1318	GCCTTCAAAAGCCCGCCCACTGGCTATTGCTTTCAGGAGCGTCTGGGGCTGGGC	1377
Db	101762	GGGATGAAGTCCGTGGCCCGCTGCCGCTCTACCCGGTGCAGGAAGTCTGGCCCTGGGC	101703
Qy	1378	CCGAGGCCCGCATGAATCTCCCGGACGGCTGGGGGACAACTGGGCGTGGCGCTACGCC	1437
Db	101702	AGCAGGCCCGGATGAACTACCCGGGAAGGCCCTCGGGGAACTGGGCCCTGGCGCTCCTC	101643
Qy	1438	GAAGCGCACTCGAGCCCGGTCTGSCCGCGGAATGCGGGGCCCTGGCGAGGCCAGCCAG	1497
Db	101642	CCGGGGAGCTTTCGCCGAGCACCGGCGAGGCTTAGGGCCATCGCCGAGGCCACGGAA	101583
Qy	1498	CG 1499	
Db	101582	CG 101581	

RESULT 9  
AP008226 12/c  
WPCOMMENT

Sequence Split	Into 15 Fragments	Begin	End
AP008226_00	1	100001	110
AP008226_01	100001	210	110
AP008226_02	20001	310	210
AP008226_03	30001	410	310
AP008226_04	40001	510	410
AP008226_05	50001	610	510
AP008226_06	60001	710	610
AP008226_07	70001	810	710
AP008226_08	80001	910	810
AP008226_09	90001	1010	910
AP008226_10	100001	1110	1010
AP008226_11	110001	1210	1110
AP008226_12	120001	1310	1210
AP008226_13	130001	1410	1310
AP008226_14	140001	1510	1410
AP008226_15	150001	1610	1510
AP008226_16	160001	1710	1610
AP008226_17	170001	1810	1710
AP008226_18	180001	1849	1810

Continuation of 13 of 19) of AP008226 from base 1200001 (AP008226 Thermus thermophilus H. AP008226 18 1800001 1849742)

Query Match	42.4%;	Score 636.8;	DB 1;	Length 110000;
Best Local Similarity	65.0%;	Pred. No. 1.1e-97;		
Matches 976;	Conservative	0;	Mismatches 517;	Indels 9;
Gaps	2;			

QY 1 ATGCAACTCCAAACGCGCTTTTGGAAATTTGCTCCACCCACACAGTTTTCGGGTCGCTGG 60  
 Db 3076 ATGAGCTTCCCGCGCTTTTGGTCTGCTTCTCCACCCACAGCCTCCCGGCCCTAC 3017  
 QY 61 GGAATTGGGGTCTGGGCGGAGCGCGGTTTGGATGCTGGCTGGCGGATGCGGGA 120  
 Db 3016 GCGTTCGGCGCTTGGGCGGAGGCGCGGACTTCTCTCGCTTCTCAAGAGGAGCGAGG 2957  
 QY 121 GCCCGCTGGTGGAGGCTTACCGCTGGCGCTTACAGTGTACGGCGACTCGCGTACCAG 180  
 Db 2956 GGGGGTACTGGGAGGCTTCTCCCTTGGGCGGCGGCGGCTATGGCGACTCCCGCTACAG- 2897  
 QY 181 TCCCTTCTCGGCTTTTGGCGGTAAACCGCTATTGTTGACCCCGAGATGCTGATTGAAAAA 240  
 Db 2896 TCCCTTACGCGCTTTCGCGGAAACCCCTACCTCATAGACCTTACGCGCCCTCGCGGAAAG 2837  
 QY 241 GCGTGGCTGGAACMAAGCGAAGCGCCCGCGGTATCCGACCCAGCGCGGTGATATGGC 300  
 Db 2836 GGCCTACGTG-----CGCCTGGAGGACCCCGGCTTCCCGGCGGCTTCCGAGGAGGCTACCGC 2783  
 QY 301 TGGCTTTACAGACCGCTGCGCCCTGCTGGCGGCGCTTTCGGGCGGCTTTCGGGCGAAG 360  
 Db 2782 CTCCTTACCGCTTGAAGTGGCGCGCTTGAAGAGGCGCTTTCGGGCGCTTCAAGGAAAG 2723  
 QY 361 GCTTCGCGCCAGGATAGACCCGACTGGAAGCCTTTATCGAGGCGGAGCGCTTCTGGCTG 420  
 Db 2722 GCTTCCCGGAGGCGGAGGCTTTCGCGCTTTCGGGAGGAGGCGCTTGGTGGCTC 2663  
 QY 421 GAAGACTATGCGCTTTTATGGGCTTCAAGACCGGTTTGAACCGCAAGCGCTTGAACGAG 480  
 Db 2662 GAGGACTATCGCCCTTTCATGCGGCTTGAAGGCGCGCACGGGCGGCTTTCGCTGAACCGG 2603  
 QY 481 TGGAGCCCGAGCTGGCGGAGCGTGAACCGCTGCGCTGCGGCGGCGCTTTCGGGAGCTG 540  
 Db 2602 TGGCCCTTCCCTTGGCGAAGCGGAGGAGGAGGCGCTTTCGGGAGGCGGAAAGCGCGTTG 2543  
 QY 541 GCCGAGGAGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTATCTGGAATGGGCGAG 600  
 Db 2542 GCCGAGGAGTGGCTTTCAGGCTTTCACCGAGTGGCTTCTTCCGCGAGTGGGCGCC 2483  
 QY 601 ACCAAGGCTATCGGAATCCAAAGGAGTTTCAATATCGGCGATATGCCATCTTTTGTG 660  
 Db 2482 TTGAAGCGGAGGCGGAGGCTTGGGCTATCGGATCATCGGAGCATGCCATCTTCGTG 2423  
 QY 661 GCCTTCGATCTCAGATCTGGGCAACCGCGACTTCTTACCTCGAGCGCGATGCG 720  
 Db 2422 GCCGAGACTCCCGGAGTCTGGGCGGCGGCGGAGTGGTTTCACTGGACGAGGAGGCG 2363  
 QY 721 AACCCCAAGTGTGGCGGCGCTTCGCGGAGTACTTCTCCGAAACCGGCGGAGCTTGG 780  
 Db 2362 CGCCCAAGTGTGGGCGGCGGCTTCCCGGAGTACTTCTCGAGAGCGGCGGCGCTGG 2303  
 QY 781 GCGAATCCGCTTATCGCTGGGATGTGATGAAAGGAGCAACTTTGCTGCTGTCATTTGCC 840  
 Db 2302 GCGAATCCGCTTATCGCTGGGAGCTTTTGGAGCGGAGGCTTCTCTTCGATCCGC 2243  
 QY 841 CGCATAGGAGTCCGCTCAAGAGTGGCACCTGGTGGGATGACACATCTCCCGGCTTT 900  
 Db 2242 CGTCTGGAGAGGCGCTTGGAGCTTCTTCCAGTGGTGGGAGGAGTCTTCCCGGCTTT 2183  
 QY 901 GAAGCTTCTGGAGGCTTCGTTTGGCGGCGCAATGCTGTGAGGCGGCGCTGGGCAAA 960  
 Db 2182 GAGGCTTCTGGAGATCCCGGAGTGGCGGCGGCGGCTGGGCGGCGCTGGGCAAG 2123  
 QY 961 GCCCGAGGAGAGGCTGTTTGTGCTGGGCTGGGCGGCGCAATGAGCGATGCGGCCATCAT 1020  
 Db 2122 GCCCGGCGGAGAGCTTCTTCCAGAGATCCAGAGGCTTTCGCGGAGGCTTCCCGCTC 2063  
 QY 1021 GCGGAAGCTGGGCTGATCAACCGCGAGTGGAGCTTTCGCGGAGTGGCTTTCGGGCTC 1080  
 Db 2062 GCCGAGGAGCTGGGCTGATCAACCGCGAGTGGAGGCGGCTGGCGGAGGCTTTCGGCTT 2003  
 QY 1081 CCCGGCATGAAGATTTTGCAGTTTGTCTTTT---TCCGGTGGAGGAGCAACGCTTTTGGCC 1137

Db 2002 CCAGGATGAAGTCTCTGAGTTCGCTTTTACGACGCGGATGGAAGACCCCTTCTCTCCC 1943  
 QY 1138 CACAACTACCCCGGCGGAGCAATGCTGCTGTATGAGCGGAAACCCAGCAACGACACC 1197  
 Db 1942 CACAACTACCCCTGCCACGCGCGGCTGCTGTATACCCGCGACCCAGCAACGACACC 1883  
 QY 1198 ACCCTGGGATGCTTCCGACGCGCGGAGCGCGGCGCTTTCATGCGGCGCTTACCTG 1257  
 Db 1882 ACCCTGGGCTGCTTACGCAAGGCGGCGGCGGCGGCTTTCATGCGGCGCTTACCTG 1823  
 QY 1258 GCCCGCTATGCGATCCGCTTCTTGTTCGGAATACGAGTTCGCGGCGCTTTCATGCGAGCTG 1317  
 Db 1822 GCGGACTTGGGCGGATCACCTTCCGGAAGAGGAGGCTGCGCTGCGGCGCTTGTATGACACCTG 1763  
 QY 1318 GCCTTCAAAAGCCCGGCGGAGCTGCTATTTGCTTTCGAGGAGCTGCTGGGCGCTGGGC 1377  
 Db 1762 GCGATGAAGTTCGCTGGCGCGGCTTCCGCTTACCCGCTGCGAGGAGCTTGGCGCTGGGC 1703  
 QY 1378 CCGGAGCGCGCATGAATTTCCCGGAGCGCTGCGGCGGAGCAACTGGGCGTGGCGCTACGCC 1437  
 Db 1702 AGCAGCGCGCGATGAATACCCCGGAGGCGCTTTCGCGGAGCTTGGCGCTTGGCGCTCTC 1643  
 QY 1438 GAAAGCGACTTCGAGCGCGCTTTCGCGCGGAGCTGCGGCGCTTTCGCGGCGGAGCGGCGAG 1497  
 Db 1642 CCGGCGGAGCTTTCGCGGAGCGGCGGCGGAGCTTTCGCGGCGGAGCGGCGGCGGCGGAG 1583  
 QY 1498 CG 1499  
 Db 1582 CG 1581

RESULT 10  
 AY459351  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

AY459351 1503 bp DNA linear BCT 30-NOV-2003  
 Thermus scotoductus 4-alpha-glucanotransferase gene, complete cds.  
 AY459351  
 AY459351.1 GI:38505478  
 Thermus scotoductus  
 Thermus scotoductus  
 Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
 Thermus.

REFERENCE  
 1 (bases 1 to 1503)  
 Park, K.-H. and Park, J.-H.  
 Direct Submission  
 Submitted (05-NOV-2003) Food Science & Technology, Seoul National  
 University, San 56-1, Shillim-dong, Kwanak-gu, Seoul 151-742, Korea  
 Location/Qualifiers

source  
 1..1503  
 /organism="Thermus scotoductus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:37636"  
 1..1503  
 /codon\_start=1  
 /transl\_table=11  
 /product="4-alpha-glucanotransferase"  
 /protein\_id="AA23241.1"  
 /db\_xref="GI:38505478"  
 /translation="MELPRAYGLLLHPTSLPGYGVGLGEAFALFLGESGARYW  
 QVPLPGTGYGDSYQSFSAFAGNPLYLDLRRRRGYLRLEDPPGPRGVVYGLYAW  
 WKWPLREAPRGQERAREKEAPQFCTRRSLDDYTLPMALKPTGGLPWNRP  
 MFLUREKALREAFSLDSQEVAFHTAWTFPRQWALREAAALGLSLIGDMPFV  
 AEDSAEVWAHPFHLDEBRTVAVGPPDYFSETGQRWGNPLRYEWEVGFSGF9F  
 IERLKALEFLHVIDHPRGEAYWEIPASCTAVEGRVKAPEGKLFQKIQTETGR  
 VFLAEDLVITPEVALRDRFLGKLVQAFDGMENPLPHNYPKSHGRVVVVTG  
 THDNDTLGWYRTATPHERAFLGRVLAEWGIGFORREEIPWALMHLGKMSVARLVVP  
 VQDVLASGEARMYFPGPSGNWMLRPGQLLPEHGERLRNWAETGRV"

CDS  
 1..1503  
 /transl\_table=11  
 /product="4-alpha-glucanotransferase"  
 /protein\_id="AA23241.1"  
 /db\_xref="GI:38505478"

ORIGIN  
 Query Match 39.7%; Score 596.2; DB 1; Length 1503;  
 Best Local Similarity 63.4%; Pred. No. 1.4e-90;  
 Matches 948; Conservative 0; Mismatches 538; Indels 9; Gaps 2;

QY	1	ATGCAATCCAAACGCGCTTTTGAAATTTTCTCCACCCACCAAGTCTTTCGGGTCTCTGG	60
Db	1	ATGGAGCTTCCGCGCGCTATGGCTCTCTCTGACCCACCAACGCTTCCCGAGCCCTAT	60
QY	61	GGGATTTGGGGCTCTGGGCGCGAGGCGGAGCGGTTTGGACTGGCTGGCGGATGCGGGA	120
Db	61	GGGGTTGGCGCTCTTGGGGAGGAGGCTTTTGGCTTCTCCGATCTCTGGGGAGTCTGGG	120
QY	121	GCCCGCTGGTGGCAGGTCTTACCGCTGGGCGCTTACCAATTTACCGGCACTCGCGTACCAG	180
Db	121	GCCCGCTACTGGCAGGTTTGGCTCTTAGGCGCCACAGATACGAGACTCCCGCTACCAG	180
QY	181	TCCTTCTGGCTTTTGGCGGTAAACCGTATTTGGTTGACCCCGAGATGCTGATGAAATA	240
Db	181	TCCTTCAGCGCTTTTGGCGGAAACCCCTACTCATCGACC-----TCGCGCTAAAGCGG	234
QY	241	GGCTGGCTGGAACAAAGCGAAGCGCCCGCGGTATCCGACCCAGCGCGTGGATTATGGC	300
Db	235	AGCGGGGGTACCTTCCCGCTTCGAGGACCCCGGTTTCCCGAGGGCGGTTGATACGGC	294
QY	301	TGGCTTTACCAAGACCCCGCTGTGGCGGCGCTTTTCGGGGGTTTCGGGCAAGG	360
Db	295	TGACTCTACGCTTGAAATGCGCGCGCTTACGAGAGCCCTTCCGGGGTTTCCAGGAGCGG	354
QY	361	GCTTGGCGCCAGGATTAAGACCGAGCTGGAAGCCCTTTATCAGAGCGCGAGCCTTCTGGCTG	420
Db	355	GCTAGCGGGGAAAGGAGGCGCTTTTCAGCGCTTTTGGGACCCCGGGAAGAGCTGGTTG	414
QY	421	GAAGACTATCGCTCTTTATGGCCCTCAAGACCGGTTTTCAGCGCAAGCCCTGAACGAG	480
Db	415	GATGACTACACCTCTTTCATGGCTTGAAGCCACCCAGAGCGCTTCTTGGNACCGC	474
QY	481	TGAGGCCCCGAGCTGCGGACCGGTGAACCGGCTGCCCTGCGAGGGCGCTGAGGAGCTG	540
Db	475	TGSCCATGCCCCCTACGCTGAGGAGGAAAGGCTTTAAGGGAAGCATCTTTGGCCCTC	534
QY	541	GCCGAGAGGTGGCCCTTTACGAGTGAATTCAGTGGCTTTTATCTGGAATGGGCGAG	600
Db	535	TCCNAGAGGTGGCTTCCAGCGCTGACCCAGTGGTTTCTTCCGCGAGTGGCAGGCC	594
QY	601	ACCAAGCCTATCCGGAATCCAGGGGATTCAGATTATCGCGCATATGCCATCTTTGTG	660
Db	595	TTGAGGAGCGCGGAGGCTTGGGCTCTCTTGATGCGGACATGCCCACTTCCTG	654
QY	661	GCCTTCGATTCTCAGATGCTGGGCGCAACCCGAGTACTTCTACCTCGAGGCGGATGGC	720
Db	655	GCCGAGACTCCGCGAGGCTTGGGCGCCACCCCGAGTGGTTCCACCTGGATGAGGGA	714
QY	721	AACCCACGCTGGTGGCGGCTTCCGCGGACTACTTCTCGAAACCGGCGAGCTCTGG	780
Db	715	AGGCCACGCTGGTGGCGGCTTCCGCGGATTTCTTTCGGAACCGGCGAGCGCTGG	774
QY	781	GGCAATCGCTCTATCGCTGGGATGATGGAAGGACAACTTTGCTGTGCTATTGCC	840
Db	775	GGGAATCCCTGTATCCGCTGGAGGTCTGGAAGGAGGGGTCTCTCTCTGGAATAGAG	834
QY	841	CGCATAGGCACTCGCTCAAGCACTGCACTGCTGGCGCATTCGACCACTTCCCGGGTTT	900
Db	835	CGCTTAGGAAGCGCTTGGAGCTTTTCCACTGCTGGCATCGACCACTTCCGGGCTT	894
QY	901	GAAGCCTACTGGGAGTTCGTTTGGCGGCGCCAAATGCTGTGAGGGGCGCTGGGTCAA	960
Db	895	GAAGCCTACTGGGAGATCCCTGCTCTGCGCCACCGCAGTGGAGGAGCGCTGGGTGAAG	954
QY	961	GCCCCAGGAGAGCTGTTTGTCTGGCTGGCGGCCCAACTGAGCGATGGCCCATATT	1020
Db	955	GCCCCGGGGAGAGCTCTTTTCAGAGATTTCAGGAACCTTTGGCGGGTGCCCATCTG	1014
QY	1021	GCCGAAGACCTGGGGGTGATCACCCCGAGGTGGAGGCTTTCGCGATGGCTTCGGGTTC	1080
Db	1015	CGGAGGACCTGGGGGTGATCACCCCGAGGTGGAGGCTTTCGAGGAGCGGTTTGGCTT	1074
QY	1081	CCCGCATGAAGATTTTGCAATTTGCTTTTCCGGTG---AGCAACGCTTTTGGCC	1137
Db	1075	CGGGGATGAAGTCTCTGAGTTGCTTTGACGGGGGATGGAACCCCTTCTTGCC	1134
QY	1138	CACAACTACCCCGCACCGCAATTTGGTGTGTACAGCGGAACCAACGACGACAC	1197
Db	1135	CACAACTACCCCTCCACCGCGGGTGGTGTCTACACCGGCAACCAACGACGAC	1194
QY	1198	ACCTTGGATGTTCCGACCGCGCGAGGCGGAGGCGCTTTCATGCGGGCTACCTG	1257
Db	1195	ACCTTGGGCTGTACCGACCGCTACCCCGACGAGCGGCTTCTTGGCGCTACCTG	1254
QY	1258	GCCCGCTATGCGATCCGTTTGTTCGGAATACAGGTTCGCGGCGCTTTGATCGAGCTG	1317
Db	1255	GCGGATGGGGATTTGGTTCCAGAGGGAAGAGAGATCCCTTGGGCGCTCATGCACTG	1314
QY	1318	GCCTTCAAAAGCCCGCCCAAGCTGCTATTTGCTTTCGAGACCTGCTGGGCTGGG	1377
Db	1315	GGGATGAAGTCCGCTGGCCAGGCTGGCCGCTTACCCCGCTGAGGATGTTTGGCGTTGGGA	1374
QY	1378	CCGAGGCGCATGAACTTCCCGGACGCTGGGGGACAACTGGGCGTGGCGCTACGCC	1437
Db	1375	AGGAGGCGCGGATGAATACCTTGGGCGGCGCTTCCGCGCACTGGGCTTGGCGCTTAGG	1434
QY	1438	GAAGGCGACTTCAGCGCGCTCTGGCGCGGGACTGCGGCGCTTGGCGGAGGCA	1492
Db	1435	CCCGGCGAGCTCTGCTCCCGGAGCACGCGGAACGCTTCTGCTGGATGGCGGAGGCA	1489
RESULT 11			
AE002007			
Deinococcus radiodurans R1 section 144 of 229 of the complete			
chromosome 1.			
AE002007.1 GI:6459402			
Deinococcus radiodurans R1			
Deinococcus radiodurans R1			
Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;			
Deinococcaceae; Deinococcus.			
1 (bases 1 to 10747)			
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,			
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,			
Hoffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,			
Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zaleski, C.,			
Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.			
Genome sequence of the radioresistant bacterium Deinococcus			
radiodurans R1			
Science 286 (5444), 1571-1577 (1999)			
10567266			
2 (bases 1 to 10747)			
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,			
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,			
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,			
Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zaleski, C.,			
Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,			
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,			
Smith, H.O., Venter, J.C. and Fraser, C.M.			
Direct Submission			
Submitted (08-NOV-1999) The Institute for Genomic Research, 9712			
Medical Center Dr, Rockville, MD 20850, USA			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
Source			
1. 10747			
/organism="Deinococcus radiodurans R1"			
/mol_type="genomic DNA"			
/strain="R1"			
/db_xref="taxon:243230"			
/chromosome="1"			
125..2242			
/gene="DR1635"			
125..2242			
/gene="DR1635"			
gene			
CDS			

/note="similar to GB:AL009126 percent identity: 54.60;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter, ATP-binding protein, EF-3  
family"  
/protein\_id="AAFI1195.1"  
/db\_xref="GI:6459403"  
/translation="MTAGACAWTCTACCAASGAGNAPTRSSGPARTREGERM  
LIRPSIREVSFTGCTGDRAGAVRILGPVLIALSHAKEYGLTLVLODITFAVEP  
GDRVLVNGRAGKSTLLRLTGQLPDGGEVRHAPNVRARALQQDPVFPAGATIDTV  
LESFADLDALAEALAAASGSEAILHHEALLEQPARRGYRRSRKEQVTLA  
FGRGREDHPTVSGGSTRGLAALLVENPDVLLIDEPNHLIDIVMEWELGFLGR  
YPCAVLISHDROFLONVKTAYLRGGTKVYAGGVTKFREOLDLEQOALFAQO  
SKOAILQSSADRMKTIWLGMSKLARRAKQARVDRMOKGAVAAPPERTRITFH  
APSGEVLDARHLTRMGERTIPADLDLQIRQERVAIIGRNGACKTILLRLGLD  
PSDDPRGVLTVGARVYDQALRGVDSQTLVDVAREYVQKDEARTLLGTFMFPF  
DHDKPARILSGGERARLALLQAQEDHNLVLIDEPNHLMELEAEDALAAATGT  
LVMSHDRAFIEGLADRIWLIEDQFVEYFGWEDYRAKHSAALEAAARTVAAATP  
KVALAPKAGLWHLKREVEAIEADIAELAELEGAQALAEADMDNADFTALGQAHEL  
EVQLEDMALWGEKQSEVAKG"  
complement (2246. .2614)  
/gene="DR1636"  
/genes="DR1636"  
/note="similar to GB:AL009126 percent identity: 72.34;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFI1196.1"  
/db\_xref="GI:6459404"  
/translation="MRRARQFLSLAAPPVYSRPMTAPVLEIAMLVNRPQOTAEFAAF  
AQGRUASIPGRRHELRQCEDAHKYVLLVWMDLEARTVGRGSAEYGOVRALLH  
HFYDPVVEHFVFPVTLPS"  
4271. .4828  
/gene="DR1638"  
4271. .4828  
/gene="DR1638"  
/note="similar to GB:AL009126 percent identity: 64.37;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFI1197.1"  
/db\_xref="GI:6459405"  
/translation="MKLYTRTGQGTGLYGDVRSKAPRVEAYGTVDLSALGLA  
RAHSTDALDLDLEYLQNALFDVGDALATRSQTPYEKNLVRIDAEDVAFLEAMIDRYQ  
EAAPPTGTVHPGTPAAALOLARAVARRAERDVIRLLDVEANAQVQVYLNRSVL  
LFTMAVAVARSGLSEANTVKER"  
4916. .5236  
/gene="DR1639"  
4916. .5236  
/gene="DR1639"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAFI1200.1"  
/db\_xref="GI:6459408"  
/translation="WNALVLLAPLPLALLGSAALPPSGVGVVNHSGAGVQNVRCRA  
GECWKDRLASGERMLVTSEAGQAPVRLTVGGAPATGQVTAQRGRGVQLMLERDRAA  
AGRQ"  
5301. .5492  
/gene="DR1640"  
5301. .5492  
/gene="DR1640"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAFI1201.1"  
/db\_xref="GI:6459409"

/translation="MAATAMKGLRFRPGELRRSEVPRTNLLPGAGOPARARREDRWRA  
WLLVLSVCLLALLGVTL"  
5546. .6040  
/gene="DR1641"  
5546. .6040  
/gene="DR1641"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAFI1202.1"  
/db\_xref="GI:6459410"  
/translation="MTDRSDRQLAFARLLPRLFRGQYVGVTEATLSDLSBQAQAH  
AGCLPHSVAEALLAHVNNWRMLDIEMQCALPYPPHAAADTWPOVGAGDWGNVKNFY  
ELLARIDTHAARPDLASPVNHEETIGELLADFALHTAHFGQIISVRQALGAWPPAGG  
GUTW"  
6037. .6528  
/gene="DR1642"  
6037. .6528  
/gene="DR1642"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAFI1203.1"  
/db\_xref="GI:6459411"  
/translation="MSEOPADQFANAAOLYLGSVANTWTERAISLSAADARRPAG  
LPHSAAQIVAHVQFWQHYLLAVIAGENPPAVAHAAEGNPAGELPAGEWELRRFLRD  
AQRLRDLTLASLTASDPKRPYAVALTNSAGHSVYHLGVVTVRQALGLWPVVGCG  
ATW"  
6522. .6767  
/gene="DR1643"  
6522. .6767  
/gene="DR1643"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAFI1204.1"  
/db\_xref="GI:6459412"  
/translation="MVSVELNAGNSSRASHVWRVDDQGRSEIYRRSWTTPEVSAPM  
LGLQLFGVDPDLHATESAYGFWRELGVTVTPQTTA"  
6740. .7372  
/gene="DR1644"  
6740. .7372  
/gene="DR1644"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAFI1205.1"  
/db\_xref="GI:6459413"  
/translation="MDRAADYGLTFEFDAPALRVYVSGDPNGRTFAEADTLLGRQI  
AQVHQAAASFQGVSGQRQPLEQFYRALETVRATAPYAPQNWAGHWDTVYRFAA  
VPPRQAAPMLLDWNEQFVWRGQPYALVDVEASATAPPELDLTFWELLIPAGAPAQ  
AFQGYRVRPMDLNPHRACELILLALESGSTRDAAQMLAQPALVETA"  
complement (7382. .8128)  
/gene="DR1645"  
complement (7382. .8128)  
/gene="DR1645"  
/note="similar to GB:M95047 SP:E37457 GB:X52093 PID:153897  
percent identity: 56.50; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-N-acetyl-D-mannosaminuronic acid  
transferase, putative"  
/protein\_id="AAFI1198.1"  
/db\_xref="GI:6459406"  
/translation="MSLSKSLPERLVFLDLPDVTLDALDRLSGLWLPESPTPTTVV  
TLNPEFVQSRTQPEFVRAMQEDLVTDGVGIIVYARQLYCAEYVPAFGFDLVKGLM  
ERHGPRLRVFLGSKPGVAQNAQNAVDYGIQVAGVHHGFGPDEDQORVAVELVGSAG  
AHLTLTGAGRQEIFNAYWQIHRAPVAIGCGGVIDVLGTAQLAPEWTRKGVWVW

gene  
CDS  
8134..8913  
/gene="DR1646"  
8134..8913  
/gene="DR1646"  
/notes="similar to GB:AL123456 percent identity: 49.31;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="nitrogen regulator, putative"  
/protein\_id="AAPI1199.1"  
/db\_xref="GI:6459407"

Query Match 32.8%; Score 492.8; DB 1; Length 10747;  
Best Local Similarity 60.5%; Pred. No. 2.6e-73;  
Matches 908; Conservative 0; Mismatches 567; Indels 25; Gaps 5;

QY	1	ATGCACTCCAAACCGCTTTTGGATTTTCTCCACCCACCGCTTTTCCGGTTCGGTGG	60
DB	2680	ATGACACTTCAACGTTCCAGCGCGTCTGCTGCACCCACCGCTTGCCTCGACCTTAT	2739
QY	61	GGGATTGGGCTCTGGGCGCGAGCGGTTTGGACTGGCTGGCGGATGGGGGA	120
DB	2740	GGCATCGCGAATCTCGCGGCTCTGGCGGCACTTCGTGACTGCTCGCGAGCGGG	2799
QY	121	GCCTGCTGGTGGAGTCTTACCGCTGGGCGCTACCACTTACCGCGACTCGCGTACCAG	180
DB	2800	CAGCGCTACTGGCAGTCTGCGCTCGGCGCCACCGCTTACGGCGACAGCCCTATCAG	2859
QY	181	TCCTTCTCGGTTTTCGGGTAAACCGTATTTGGTTGACCCCGAGATGCTGATGAAAA	240
DB	2860	GCCTTTTCGGGTTTCGGCGCAACCTTACCTATCGACCTGACGACGCTGAGGGAAGAA	2919
QY	241	GGCTGGCTGGAACAAA-----GGGAAGCGCCCGCGCTATCGACCGCGGCTGGAT	294
DB	2920	GGGCTGTGACGACCGGATTTGAGCTATGCGGCTTCAACCGCGCGGTTGAT	2979
QY	295	TATGGCTGGCTTTACCAAGACCGCTGCGCCCTGTGTGGCGGGCTTTTCGGGGTTTCGG	354
DB	2980	TTCGGACCGAGTTTGTGTGGCGCAACAGATGCTCGACCGCGCTACGCGCACTTTATT	3039
QY	355	GCAAGGCTTCGGCGCCAGGATAGAACCGGACTGGAAGCTTTATCGAGGCGGAGCTTC	414
DB	3040	TACGGCGAACACGCGCAACTGACGGCGCGTTCGGGCGCTTCAAGGCGCGAGAGCGGC	3099
QY	415	TGGCTGCAAGACTATGCTCTTTATGCGCTCAAGACCGGTTTGAACGCAAGCCCTGG	474
DB	3100	TGGCTGCAAGACTATGCTCTTTATGCGCTCAAGACCGGTTTGAACGCAAGCCCTGG	3159
QY	475	AACGAGTGGAGCCCGAGCTGCGGACCGTGAACCGGCTGCGCTGGCGCAGGCGCGTGAG	534
DB	3160	AACGCTTGGAGCAGGCGCTGCGGACCGGACCGCGAGCGGCTCGCTCGCGCCGGAM	3219
QY	535	GAGTGGCGGAGGAGTGGCCCTTACAGTGGATTCAGTGGCTTTTATCTGGAATGG	594
DB	3220	G-GTTGGAAGCGCCATCAACCGGTGCGATTCTGTGCACTTCTCTTTTTCGGCGAGTG	3278
QY	595	GGCAGACCAAGGCTATCGCAATCCAGGGATTCAGATATCGGCGATATGCCCATC	654
DB	3279	ACAGCGTGTGTGATGATGAGCGCGGCTGAGCGCGGCTGAGATATTCGGGACATTTCCAT	3338
QY	655	TTTGTGGCTTCGATTTCTCAGATGTCTGGGCGCAACCGCAGTACTTCTACTCGAGGCC	714
DB	3339	TTTGTGGCTTCGATTCAGCGACGCTTGGGCGCAACCGGAGCGATTTCTACTCGAGAC	3398
QY	715	GATGGCAACCCACCGGTGGGCGGCTTCGGCGGACTACTTCTCGAAACCGGCGAG	774
DB	3399	CAGGCGCAGCGAGTGGTTCGGGCGTGGCGCGGACTATTTCTCGGAGCCGGCAA	3458
QY	775	CTCTGGGCAATCGCTCTATCGCTGGATGTGATGGAAGGGAACAATTTCCTGGTGC	834
DB	3459	CTGTGGGCAATCCCTCTTACACTGGCGCCCAAGATGAAGAAGACGGCTTCGGCTGTGG	3518

Sequence split into 39 fragments LOCUS AE017180 Accession AE017180

Fragment Name	Begin	End
AE017180_00	1	110000
AE017180_01	100001	210000
AE017180_02	200001	310000
AE017180_03	300001	410000
AE017180_04	400001	510000
AE017180_05	500001	610000
AE017180_06	600001	710000
AE017180_07	700001	810000
AE017180_08	800001	910000
AE017180_09	900001	1010000
AE017180_10	1000001	1110000
AE017180_11	1100001	1210000
AE017180_12	1200001	1310000
AE017180_13	1300001	1410000
AE017180_14	1400001	1510000
AE017180_15	1500001	1610000
AE017180_16	1600001	1710000
AE017180_17	1700001	1810000
AE017180_18	1800001	1910000
AE017180_19	1900001	2010000
AE017180_20	2000001	2110000
AE017180_21	2100001	2210000
AE017180_22	2200001	2310000

RESULT 12  
AE017180.12  
WPCOMMENT

AE017180\_23 2300001 2410000  
AE017180\_24 2400001 2510000  
AE017180\_25 2500001 2610000  
AE017180\_26 2600001 2710000  
AE017180\_27 2700001 2810000  
AE017180\_28 2800001 2910000  
AE017180\_29 2900001 3010000  
AE017180\_30 3000001 3110000  
AE017180\_31 3100001 3210000  
AE017180\_32 3200001 3310000  
AE017180\_33 3300001 3410000  
AE017180\_34 3400001 3510000  
AE017180\_35 3500001 3610000  
AE017180\_36 3600001 3710000  
AE017180\_37 3700001 3810000  
AE017180\_38 3800001 3814139

Continuation (13 of 39) of AE017180 from base 1200001 (AE017180 Geobacter sulfurreducens)

Query Match 28.1%; Score 421.8; DB 1; Length 110000;  
Best Local Similarity 58.0%; Pred. No. 1.7e-61;  
Matches 838; Conservative 0; Mismatches 577; Indels 30; Gaps 4;

Qy 1 ATGCAATCCAAACCGCTTTTGGAAATTTGCTCCACCCACCAAGTTTTCGGGTGCTGG 60  
Db 88728 ATGAGACAGAAACGTCGACGGGATCTGCTCCATCCGACTTCCCTGCCGGTCCGGGG 88787

Qy 61 GGGATTGGGCTCTGGCCCGGAGGCGGAGGGTTTGTGACTGGCTGGCGATGCGGGA 120  
Db 88788 GGGATCGGATCGCTCGGGAAGAGTGCCTCATTTTCATCGATTTCCTCGAAGAGGGCGGG 88847

Qy 121 GCCCGCTGTGGCAGGCTTACCGCTGGCCCTACCAAGTTACGGCGACTCGCGTACCAG 180  
Db 88848 CAGAGCTCTGGCAGGTTCTTCGCTGGGCGCGCAGCCTACGGCAATCCCCATATCC 88907

Qy 181 TCCTTCTCGCTTTTTCGGTAACCCCGTATTTGGTTGACCCCGAGATGCTGATGAAAA 240  
Db 88908 TGCTATTTCGGCTTTTGTGCAACCCCTCTGATCGACTGGCGACCTCTGTGGATGAG 88967

Qy 241 GGTGGCTGGAACAAAGCGAAGCGCCCGCGTATCCGACCGCGCGTGGATATGGC 300  
Db 88968 GGGGACCTGGAGCCGAGGAGGCTGTGGCGAGGTGTGCGCGATCATGTGAGCTTTCCG 89027

Qy 301 TGGCTTTACCAAGCCGCTGCGCCCTGCTGGCGGGCTTTTCGGGGTTTTCGGCAAG 360  
Db 89028 GCTGTGATCGGTTCAAGTCACGCGCTCTCGGACCGCTCGGCGCGGTTTCCAGCGCCAC 89087

Qy 361 GCTTCGCGCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGCCGAGCGCTTCTGGCTG 420  
Db 89088 GGTGACACGATCGTAAACAGGAGTTCTGGCACTTTTTCGATACTACCCC---CTGGCTT 89144

Qy 421 GAAGACTATGCGCTCTTTATGGGCTCAAGACCGGGTTTTCAGCGCAAGCCCTGGAACGAG 480  
Db 89145 CACGAGCTGCGCTGTTTCATGGCTCTCAAGAGCAATTTTCGGCGCAAAAGCTGGAATACC 89204

Qy 481 TGGAGCCCGAGCTGCGGACCGTGAACCGGCTGCTGCGAGGCGCCGTGAGGAGCTG 540  
Db 89205 TGCGCCCAAGGAGATTGCGCCCGGGAGTTCGGCGCTCGGCGCTCGGCTTCGGTTTCGGCTG 89264

Qy 541 GCCGAGGAGTGGCCCTTTACGAGTGGATTTCAGTGGCTTTTTCATCTGGAATGGGCGCAG 600  
Db 89265 GGGTGGCCATTGGCGNAACAAGTATATCCATGGCAGTTTTCGGTCACTGCGGACAT 89324

Qy 601 ACCAAGCCCTATCCGAATCCAAAGGGATTCAGATTATCGCGATATGCCCATCTTTTGTG 660  
Db 89325 GTCAGGGAATATGCCAACAGAGGGAATTCGGTGTGGCGATATCCCATCTTTTGTG 89384

Qy 661 GCCTTCGATTCCTCAGATGCTGGCCCAACCGCAGTACTTCTACCTCGAGGCGGATGCG 720  
Db 89385 GCTTTTCGATTTCAACCGGATGCTGGAGCAATTCGGGCTTTTCAAGCTTCGACCAAGGGG 89444

Qy 721 AACCCCAAGCTGTGGCGGCTTTCGGGACTACTTCTCCGAACCGCAGCTCTGG 780  
Db 89445 AAGCCCAAGCTGTGGCGGCTTTCGGGACTACTTTCAGCAAGACAGCGCGCATGG 89504

Qy 781 GGCAATCCGCTCTATCGCTGGATGTGATGAAAGGACAACTTTTGCCTGCTGATTCGCC 840  
Db 89505 GGGAAACCTCTCTATGATCGGATGCCATGCCGCCGATGATTTTCGCTGGTGGATCAG 89564

Qy 841 CGCATAGGAGTCGCTCAAGCAGTGCACCTGCTGGTGCATCGACCACTTCCCGGGGTTT 900  
Db 89565 CGGTTCCGGGTTTCGTTTCGGCTCCATGATGTGTGAGGTTCGATCACTTCCGAGGGTTC 89624

Qy 901 GAAGCCTACTGGGAGGTTTCGTTTGGCCGCCCAATGCTGTGAGGGGCGCTGGGTCAA 960  
Db 89625 GAGCGTGTCTGGGAGGTTTCGGCCAAAGAAAGACGCGCGTGAACGCGCAATGGGTCCCC 89684

Qy 961 GCCCAGGGGAGAGGCTGTTTTCGCTGGGTGCGGSCCAACTGAGCGATGCGCCCATATT 1020  
Db 89685 GCTCCGGGAGCGGACTCTTTGATGCGGTTATCGGGCGGTGGGTCACTGCCGATCAAT 89744

Qy 1021 GCCGAAGACCTGGGGGTGATCAACCCCGAGGTGGAGGCTTTGGCGATGGCTTCGGGTTT 1080  
Db 89745 GCCGAGGACCTGGGAGTCAATACCCCGCGCTGGAGGAACCTGCGGACCGGTACCGTTT 89804

Qy 1081 CCCGGCATGAAGATTTTTCAGTTTTCGTTT---TTCGGTGGAGACAAACGCTTTTTCGCC 1137  
Db 89805 CCCGGCATGAAGATCTCTCAGTTTCGCTTTGATTCGGGCGCGGCAATCCCTACCTGCC 89864

Qy 1138 CAAACTACCCCGCGCACCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1197  
Db 89865 CAAACTTTCACCCGCTGACTGC-----GTGGTGTACACCGGACCCACGACAAATGATACC 89918

Qy 1198 ACCCTGGGATGTTTCGCGACCGCGGAGCGCGGCTTTCATGCGGGCTACCTG 1257  
Db 89919 ACCGAGGCTGGTTTCAGAGCTTTTCCCGCGGAGCGAGAGTTCCTGGCTATACC 89978

Qy 1258 GCCCGTATGGCATTCGTTTGTTCGGAATACGAGTTCGCGGCGCTTTTGTATCGAGCTG 1317  
Db 89979 GACACACCGGTA-----AGGAGTCCATTGGGAGCTGATCCGCTC 90020

Qy 1318 GCCTTCAAAGCCCGCGCAAGCTGCTATTTGTCCTTTGACGACGCTGCTGGGCTGGC 1377  
Db 90021 GGCATGGCTCGGTGGCGGATATGGCGATTTTCCCGCTTCAGGATGTGATGGGACTTGAT 90080

Qy 1378 CCGAGGCGCGCATGAATCTTCCCGGACGCTGCGGAGCAAACTGGGCGTGGCGCTACGCC 1437  
Db 90081 GGGTCAAGCGGATGATCTCCCGGACACCGCGGAGTGGTCTGGGCTTATAGT 90140

Qy 1438 GAAGG 1442  
Db 90141 GAGGG 90145

Sequence split into 47 fragments LOCUS BA000045 Accession BA000045  
Fragment Name Begin End  
BA000045\_00 1 110000  
BA000045\_01 100001 210000  
BA000045\_02 200001 310000  
BA000045\_03 300001 410000  
BA000045\_04 400001 510000  
BA000045\_05 500001 610000  
BA000045\_06 600001 710000  
BA000045\_07 700001 810000  
BA000045\_08 800001 910000  
BA000045\_09 900001 1010000  
BA000045\_10 1000001 1110000  
BA000045\_11 1100001 1210000  
BA000045\_12 1200001 1310000  
BA000045\_13 1300001 1410000  
BA000045\_14 1400001 1510000  
BA000045\_15 1500001 1610000  
BA000045\_16 1600001 1710000  
BA000045\_17 1700001 1810000

RESULT 13  
BA000045.35/c  
WPCOMMENT



RESULT 14	
AP008231_05/c	
WPCOMMENT	
Sequence split into 27 fragments	LOCUS AP008231 Accession AP008231
Fragment Name	Begin End
AP008231_00	1 110000
AP008231_01	100001 210000
AP008231_02	200001 310000
AP008231_03	300001 410000
AP008231_04	400001 510000

AP008231\_05 500001 610000  
AP008231\_06 600001 710000  
AP008231\_07 700001 810000  
AP008231\_08 800001 910000  
AP008231\_09 900001 1010000  
AP008231\_10 1000001 1110000  
AP008231\_11 1100001 1210000  
AP008231\_12 1200001 1310000  
AP008231\_13 1300001 1410000  
AP008231\_14 1400001 1510000  
AP008231\_15 1500001 1610000  
AP008231\_16 1600001 1710000  
AP008231\_17 1700001 1810000  
AP008231\_18 1800001 1910000  
AP008231\_19 1900001 2010000  
AP008231\_20 2000001 2110000  
AP008231\_21 2100001 2210000  
AP008231\_22 2200001 2310000  
AP008231\_23 2300001 2410000  
AP008231\_24 2400001 2510000  
AP008231\_25 2500001 2610000  
AP008231\_26 2600001 2696255  
Continuation (6 of 27) of AP008231 from base 500001 (AP008231 Synchococcus elongatus pQ)

Query Match 26.8%; Score 403.2; DB 1; Length 110000;  
Best Local Similarity 56.1%; Pred. No. 2.2e-58;  
Matches 834; Conservative 0; Mismatches 628; Indels 24; Gaps 3;

QY 12 AGCGGCTTTTGAATTTGCTCCACCCACACAGTTTTCGGGTCGCTGGGGGATTTGGGGC 71  
DB 87497 ACAGCGCGCGGATTTTGTATCCACTTCGTTGCCGGGCAATCGGACGGCGCA 87438  
QY 72 TCTGGGCGGAGCGCGAGCGGTTTGTGACTGGCTGGCGGATGCGGAGCCCGCTGGTG 131  
DB 87437 CTTTGTGGGCTTCGCGGAGTTTCTGACTGGTTGGCAACCGCGGCAACAACCTGTG 87378  
QY 132 GCAGGTCTTACGCTGGGCTTACCACTTACGCGGCTGCGCTTACCACTTCTCGCG 191  
DB 87377 GCAGGTCTTACGCTGGGCTTACCACTTACGCGGCTGCGCTTACCACTTCTCGCG 87318  
QY 192 TTTTGGGCTTAACTGCTTATTTGTTGACCCCGAGATGCTGATTTGAAAAAGGCTGGCTGA 251  
DB 87317 CTGTGGCTGGCAATCCCGCTCTGATCAGCCCTGAACTTTGGCAGAGATGGCTGGCTCA 87258  
QY 252 ACAA-----AGCGAAGCGCCCGCGCTATCCGACCCAGCGCTGGATATGCTGGCT 305  
DB 87257 AGAATCGGACTGGGCGAGCTGCTGCTTTTCCGAGCGATCGCGTTCGATTTTCCAGCGT 87198  
QY 306 TTACACAGACCGCTGGGCTTGTGGGCGGCTTTTCCGCGGCTTTCGGCGAGGCTTC 365  
DB 87197 CTGTGGCTTACGCTGCTGCGGCTTACCTTACGCGGCTTTCGCGAGGCTTC 87138  
QY 366 GGCCAGGATAGACCCGAGCTGGAAGCTTTTATCGAGGCGCGAGCGCTTCTGGCTGGAAGA 425  
DB 87137 TTTCCAGGATCGCAACTCTTTCAAGCTTCTGTGAACAGGAGCCATTTGGCTGATGA 87078  
QY 426 CTATGGCTCTTTATGGGCTCTAAGACCCGCTTTTGAAGCGCAAGCCCTGGAACGAGTGGAG 485  
DB 87077 CTACGCGCTTCTATGGGCTTAAAGCTGAGCAAGCTTACGCTTGGACAGATGGCC 87018  
QY 486 CCCCGAGCTGGGAGCGGCTGACCGGCTGCTGCGGCGGCGGCTTGGAGGCTGGCGCA 545  
DB 87017 GGAAGCGCTGCTACGCGCAACCTCAAGCTTGGCTTAAAGCCCGGAGTGGCTGGGCGG 86958  
QY 546 GGAGGTGGGCTTTTACAGTGGATTCAGTGGCTTTTATCTGGAATGGGCGCAGACCAA 605  
DB 86957 CGAATTTGCTTCCAGCAGTTTCTGAGTGGCAATTTCCGAGCAGTGGTTGGCCCTGGC 86898  
QY 606 GGCCTATGCGGAATCAAGGGGATTCAGATTTACGGCGATATGCCCATTTTGTGGCTTT 665  
DB 86897 GGAAGAGCCCAAGCCCGCATATTTCTGCTGATTTGGCGATATTCGATCTACGTGCTCA 86838  
QY 666 CGATTCCTCAGATGTCTGGGCGCAACCCGCGAGTACTTCTACTCTGGAGGCGCA---TGGCA 722

DB 85837 TGACAGTCGGACGTTTGGGCGCAATCTCAGTCTTTTGGCCCTCGATCTCGAAGACGGGCGC 86778  
QY 723 CCCACGGTGGTGGCGGGGCTTCCGCGGAGCTACTTCTCCGAAACCGGCGAGCTCTGGGG 782  
DB 86777 AGTTGATCAGCAGCGCGGTGTGCGGCTGACTATTTCTCGAAACCGGCGCACTCTGGGG 86718  
QY 783 CAATCCGCTCTATCGCTGGGATGTGATGAAAGGAGCAACTTTGCTGGTGGCATTTGCCCG 842.  
DB 85717 CAATCCGCTCTAACAATGGGCTGCGCTGAGGCGGATGCTATCGCTGGTGGTTCACG 86658  
QY 843 CATAAGCAGTGCCTCAAGCAGTGCCACCTGGTGGCGCATCGAACCTTCCGCGGGTTGA 902  
DB 86657/GCTGCAACAGCTCCTCAGCTTAGTGGACTACATTGCGCATCGACCACTTCGCGGTTTGA 86598  
QY 903 AGCTTACTGGAGGTTCCGTTTGGCGGCGCCCAATGTGTGGAGGGGCGCTGGGTCAAAGC 962  
DB 86597 GCGTCTTGGTGGTTCGCGCTTGGTGAAGAAACCGGATCGACGAGAGTGGGTCAAAGC 86538  
QY 963 CCCAGGGAGAGCTGTTTGTCTGGGTGGGCGCCCAACTGAGCGATGCGCCCATCTTC 1022  
DB 86537 CCAGGCGCTGATCTGCTGAGCAGCATTCGCAAAACCTGGGAGCGCTACCGATCTGGC 86478  
QY 1023 CGAAGACCTGGGCGTGCATCAACCCCGAGGTGGAGGCTTTTGGCGCATGGCTTCGGGTTCCC 1082  
DB 86477 AGAGGATCTCGTGTGATTTACGCGGAGGTGGAGCGCTGGCGATCGCTTTGAGCTGCC 86418  
QY 1083 CGGCATGAAGATTTTGCAGTTTGTCTTTTCCGGTGAAGCAACGCTTTTGGCCCCCA 1142  
DB 86417 GGGCATGAAGATTTTGCAGTTTGTCTTGGGCGCGCAATGCTTATCTACCGCA 86358  
QY 1143 CTACCCCGGCGACGCAATGTGTGTGTATCAGCGGAACCCAGCAACGACACCACTT 1202  
DB 86357 CAACTACTGGGTGCTGCTGGGTGGCTTTACACCGGCACTCCACGCAATGACACGACCGT 86298  
QY 1203 GGGATGTTTCCGACCGCGCGAGCGCGGCTTTCATGCGGCGCTTACCTGGGCGG 1262  
DB 86297 CGGCTGTTTCTGCTCCGCAATGACAGCATCGCCAAACGGTGTGATTTCTGGGCGC 86238  
QY 1263 CTATGGCATTCGTTTGTTCGGAATACAGAGTTCGGGCGGCTTTGATCGAGCTGGCTTT 1322  
DB 86237 AGA-----GTGGGCTGGGAATTCAGTGAAGCTGATCCGCTTGGCTTG 86193  
QY 1323 CAAAAGCCCGCAAGCTGGCTATTTGTGCTTTGAGGAGCTGCTGGGCTGGGCCCCGA 1382  
DB 86192 GAGCTCGACGGCAGATTTGGGCGATCGCACCGCTTCAAGATGTCTTGGGCTGGATAGCAG 86133  
QY 1383 GGGCCGCTATGAATTTCCCGGACGCTGGGGGACAACTGGGCGTGGCGCTTACGCCGAAG 1442  
DB 86132 CGCCCGCATGATTCGACCGGGCAAGCCACCGCACTGGGACTGGGCTTCACTGCTCCGCA 86073  
QY 1443 CGACTCGAGCCGCTTGGGCGCGGAGCTGGGCGCTTGGGCGCTTGGGCGAG 1488  
DB 86072 CTGGCTGACGGCGATCGTGGCCCAACGCTGCGGCGACTCTCGCAG 86027

RESULT 15

BA000022\_02/c

WPCOMMENT

Sequence split into 36 fragments LOCUS BA000022 Accession BA000022

Fragment Name	Begin	End
BA000022_00	1	110000
BA000022_01	100001	210000
BA000022_02	200001	310000
BA000022_03	300001	410000
BA000022_04	400001	510000
BA000022_05	500001	610000
BA000022_06	600001	710000
BA000022_07	700001	810000
BA000022_08	800001	910000
BA000022_09	900001	1010000
BA000022_10	1000001	1110000
BA000022_11	1100001	1210000



QY	655	TTTTGGCTTCTGATTCCTCAGATGCTCGGCGCAACCCGCGAGTACTTCTACCTCGAGGCC	714
Db	46359	TACGTTGCCACAGTAGCGCTGCGTTTGGCGCAATCCAGAGAACCTTTTGCCTTGATCCG	46300
QY	715	GA---TGGCAACCCCAACGCTGCTCGGCGGCTTCCGCGGAGTACTTCTCCGAAACCGGC	771
Db	46299	GAACCGGAGAGCGGCGCATGATGCGCGGGTACACCGGACTATTTTCTAGTGCACAGGG	46240
QY	772	CAGCTCTGGGCAATCCCGTCTATCGCTGGGATGATGGAAGGAGCAACTTTGCTCTGG	831
Db	46239	CAACTCTGGGGTAATCTCTGTGACGACTGGGAAACCTCAAGCTACGGGCTTTGCTCTGG	46180
QY	832	TGCTTTGCGCGCATTAAGGCGAGTCCGCTCAAGCAGTGCACCTGCTGGCGCATCGACCACTTC	891
Db	46179	TGATTTAAGCGTTTAAAGCCATCTGCAATATCTAGACATTTGTCGCAATGACCACTTC	46120
QY	892	CGCGGTTTGAAGCCTTACTGGAGGTTCCGTTTGGCGGCGCCCAATCTGTGGAGGGCGC	951
Db	46119	CGGGCTTCGAGTCTCTTATTGGGAGTGCACCAAGGGGAAACCTCTGAAAATGGCGAA	46060
QY	952	TGGTCAAGAGCCCGAGGAGAGCTGTTTGTCTGCGTGGGCGCCCAACTGAGCG---AT	1008
Db	46059	TGCTATCCAGCCCGCGGCAAGGAAATTTTCCAGCCCTAGGGAAGCCCTGGGTGATAAT	46000
QY	1009	GCGCCCATCATTTGCCGAGACCTTGGGGGTGATCACCCCGAGGTGAGGCTTTTCGCGCAT	1068
Db	45999	TTACCATTTGTGCGCCAGATTTTGGGGGTAAATTAATCTCCGAGTGGAGCGCTACGGAT	45940
QY	1069	GGCTTCGGGTTCCCGGCGATGAAGATTTTGCAGTGTGCTTTTTCGGTGAAGCAACGCC	1128
Db	45939	GAATTTAACTTTCCCGGCGATGAAGTGTCTTTCGCTTTCGACTCCGACCGGGGTAAAT	45880
QY	1129	TTTTTCCCGCAACATACCCCGGCGAGGAGTGTGCTGTGCTGACAGCGGAACCCAGAC	1188
Db	45879	CCCTTTCTGCTTCAACTACAGCAATGGCAACGCGGTGCTTTTATACCGGCAACCCAGC	45820
QY	1189	AACGACACCACTCGGATGTTTCCGACCGCGCGAGGCGCGGCGCTTTCATGCGG	1248
Db	45819	AACGACACCACTCGGCTGTTTTCAGGAACGCTCAGAGGATGATCAGC-----	45771
QY	1249	GCCTACTGCGCGCTATGCGCATTCGTTTGTTCGGAATACAGGTTCGCGGGCGCTTG	1308
Db	45770	-----AAGGTGATCAATTAACCTCGCTGTGTGTCGCAACGAAGGTATCCATTGGAGCTTA	45715
QY	1309	ATCGAGCTGGCTTTCAAAAGCCCGGCGCAAGCTGGCTATTTGCTGCTTTGCAAGGACGCTG	1368
Db	45714	ATTGCGCTAGCGTCCAGTTTCGGTGGGCGGCTTAGCAATTTTCCCTCCAGGACATCCTC	45655
QY	1369	GGCTGGGCGCGGCGCGCATGAACTTCCCGGAGCGGCTGGGGGACAACTGGGCGTGG	1428
Db	45654	GGTTGGGTAGTAGCTGTGCGCATGAATTAACCGGCGACTCGCGGGGCACTGGGGTGG	45595
QY	1429	CGCTAGCGGAGGCGCACTTCGAGCGCGGCTCTGGCGCGGAGCTGGGGCGCTGGCGGAG	1488
Db	45594	CGTTACATCCCGACCACTCAAGATTTGGCTATCGGGGCAATTTAGTTTCAITACGAG	45535

Search completed: January 14, 2006, 13:11:42  
Job time : 7432 secs

BA000022_12	120001	1310000
BA000022_13	130001	1410000
BA000022_14	140001	1510000
BA000022_15	150001	1610000
BA000022_16	160001	1710000
BA000022_17	170001	1810000
BA000022_18	180001	1910000
BA000022_19	190001	2010000
BA000022_20	200001	2110000
BA000022_21	210001	2210000
BA000022_22	220001	2310000
BA000022_23	230001	2410000
BA000022_24	240001	2510000
BA000022_25	250001	2610000
BA000022_26	260001	2710000
BA000022_27	270001	2810000
BA000022_28	280001	2910000
BA000022_29	290001	3010000
BA000022_30	300001	3110000
BA000022_31	310001	3210000
BA000022_32	320001	3310000
BA000022_33	330001	3410000
BA000022_34	340001	3510000
BA000022_35	350001	3573470
Continuation (3 of 36) of BA000022 from base 200001 (BA000022 Synechocystis sp. PCC 6803)		
Query Match 21.9%; Score 329.2; DB 1; Length 110000;		
Best Local Similarity 53.3%; Pred. No. 6.4e-46;		
Matches 800; Conservative 0; Mismatches 673; Indels 27; Gaps 4;		
QY	1	ATGCACTCCAGCGCTTTTGGAAATTTTGTCTCACCCACCAAGTTTTCGGTTCGCTGG 60
Db	47019	ATGTTAGATAAAGCTGTAGCGTATTTCTGCTCATCCCACTCTTGGCCAGTCGTTT 45960
QY	61	GGGATTTGGGCTCTGGGCGGAGCGGTTTGTGACTGGCTGGCGGATGCGGGA 120
Db	46959	GGCATTTGGGACTTTGGGGAGCGTGCTTTTCAGTTTCAATTCGATTTTGGGAGCGAGAC 46900
QY	121	GCCGCTGTGTGGAGTCTTTACCGCTGGGCGCTTACCAAGTTTACGGCGACTCGCGTACCAG 180
Db	46899	CAGAGTGTATGGCAATATTTGCGGTTGGGCGGACCGGATTCGGCAATTCCTTACCTT 46840
QY	181	TCCTTCTCGGCTTTTGGCGTAACCGCTATTTGGTTGACCCGAGATGCTGATGAATAA 240
Db	46839	TGCTATTTCTGCTTAGCAATTAATCTTGTGTTGATTAGCTTGTATCGCTAGCTGAAGA 46780
QY	241	GGCTGGTGAACAAAG-----CGAGCGCCCGCGCTATCCGACCCAGCGGCTGGAT 294
Db	46779	GGATTTTGGCACCGAGTTTATAGACCAAGCCCGCCCATTTACTAACCAGGGTGGAT 46720
QY	295	TATGGCTGGCTTTTACCAAGCCGCTGCGCCCTGTTGGCGGGGCTTTTCGGGGTTTCGG 354
Db	46719	TATGACCAAGGAGTGGCTGCTAATCCAGTTTAAACAGGCGCTTTGCGGCTTCGCT 46660
QY	355	GCAAGGCTTTGGGCGGAGTAAGACCGGCTGGAAGCTTTTATCGAGGCGGAGCGCTTC 414
Db	46659	ACCAACATAGAGCTAGCAATTTGACAGGAATTTGCAAGATTTTGCACAGCCCAAGTGA 46600
QY	415	TGGCTGGAAGACTATGGCTCTTTATGGCCCTCAAGACCCGGTTTTCAGCGCAACCCCTGG 474
Db	46599	TGGCTAGCAGATTACGCGCTTTCATGGCCATCAAGAGGCGCCCAATGAGCGGTTGG 46540
QY	475	AACGAGTGGAGCCCGGAGCTGCGGACCGTGAACCGGCTGCGCTGGCGAGGCGCGGTGAG 534
Db	46539	CATCAATGGGACAGGACATTTGCTGGCGGGAACCGGAGCCCTGAAATCTGGGGCGAT 46480
QY	535	GAGCTGGCGGAGGAGTGGCCCTTTAAGAGTGAATTCAGTGGCTTTTATCTGGAATGG 594
Db	46479	CGCTGAAACCGGAAGTTTATACCATCAATTTTGTCAATTTCTTGTGTTTTCGCAATGG 46420
QY	595	GGCCAGACCAAGGCTATGCGGATCCAGGGGATTCAGATTATCGGGATATGCCCATC 654
Db	46419	CAGGAAGTCAAGGCTTAGCTTAACCAAGCGGACATAGCCATTTTGGGGACCTTACCATC 46360

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2006, 10:18:26 ; Search time 5555 Seconds  
(without alignments)  
12659.043 Million cell updates/sec

Title: US-10-618-976-1  
Perfect score: 1503  
Sequence: 1 atgcaactccaagcgcttt.....ccgaggccagccagcgct 1503

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_est3.\*
- 4: gb\_hcc.\*
- 5: gb\_est4.\*
- 6: gb\_est5.\*
- 7: gb\_est6.\*
- 8: gb\_est7.\*
- 9: gb\_gss1.\*
- 10: gb\_gss2.\*
- 11: gb\_gss3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	156.6	10.4	1569	10	CL973535 OIFCC024
2	149	9.9	877	10	CG848343 ZMWBb031
3	145.4	9.7	719	7	CK369120 zmrws055
4	145	9.6	1875	4	CNS0A259 Arabidops
5	143	9.5	762	7	CV470446 43319.1 C
6	140.2	9.3	733	1	AJ795827
7	138.2	9.2	538	3	BI720945 1031053A0
8	132.4	8.8	1155	4	AY109140 Zea mays
9	128.2	8.5	678	7	CK767135 pam01-9ms
10	126.8	8.4	683	6	CB074139
11	125.4	8.3	646	7	COL19845
12	120	8.0	605	2	BES15486 WHE0613 B
13	117.8	7.8	548	6	CF013402 QRSel12.x
14	117.4	7.8	765	6	CF437795 EST674140
15	116.6	7.8	769	7	CV302316 67345.1 S
16	114.2	7.6	576	3	BJ175019 BJ175019
17	114.2	7.6	577	3	BJ184679 BJ184679
18	112.6	7.5	666	6	CA067878 SCQAD105
19	112.4	7.5	609	6	CD481918 atr01-9ms
20	110.4	7.3	403	3	BJ943583 BJ943583
21	108.8	7.2	518	1	AW623584 EST321529
22	107.8	7.2	832	7	CO094503 GR_Eal6X

23	107.2	7.1	471	6	CB210624
24	106.6	7.1	606	6	CD442683
25	105.2	7.0	706	2	BG600511
26	103	6.9	662	3	BJ940916
27	102.2	6.8	607	8	CK189158
28	92.8	6.2	546	6	CA797031
29	91.2	6.1	508	3	BI946157
30	87.4	5.8	301	3	BM363539
31	86.4	5.7	802	3	BJ977741
32	86.2	5.7	716	6	CF512764
33	84	5.6	582	6	CD482306
34	83	5.5	395	7	CO984423
35	81.2	5.4	626	7	CK756616
36	78.6	5.2	800	1	AI823034
37	78.2	5.2	890	2	BE055692
38	77.8	5.2	695	7	CK718248
39	77.8	5.2	714	3	BI405594
40	76.8	5.1	636	2	BG580024
41	74.4	5.0	524	1	AI725928
42	74	4.9	744	9	CC943404
43	70.8	4.7	887	7	COL127810
44	69.6	4.6	513	2	BF004941
45	66.2	4.4	314	6	CA999774

ALIGNMENTS

RESULT 1  
CL973535  
LOCUS CL973535 1569 bp DNA linear GSS 21-SEP-2004  
DEFINITION OIFCC024465 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.  
ACCESSION CL973535  
VERSION CL973535.1 GI:52401596  
KEYWORDS GSS.  
SOURCE Oryza sativa (indica cultivar-group)  
ORGANISM Oryza sativa (indica cultivar-group)  
REFERENCE 1 (bases 1 to 1569)  
AUTHORS Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. M. and Wang, J.  
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis  
JOURNAL Unpublished (2004)  
COMMENT Contact: Chen Chen  
Department of Bioinformatic  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559  
Fax: 86-10-80488676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Class: exon-trapped.  
FEATURES  
source  
1..1569  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:39946"  
/clone\_lib="Oryza sativa Express Library"  
/note="Oryza sativa exon trapped genomic sequences"

ORIGIN  
Query Match 10.4%; Score 156.6; DB 10; Length 1569;  
Best Local Similarity 50.9%; Pred.No. 6.8e-28;  
Matches 398; Conservative 0; Mismatches 381; Indels 3; Gaps 1;  
QY 415 TGCGTGAAGACTATCGCTCTTTATGGCCCTCAAGACCGGTTTACGCGACCGCTGG 474  
DB 532 TGCGTGAAGACTATCGCTCTTTATGGCTATTGACAGAGTATGATGATTCCTGG 591

QY	475	AACGAGTGGAGCCCGAGCTGCGGACCGTGAAACCGGCTGCTCGGCCAGGGCCCGTGAG	534
DB	592	TATGAGTGGCCTGAAACCTCTGAAAAAACCGACATCTCGGGCCCTTGGAAAGACATATATCAA	651
QY	535	GAGCTGGCGAGAGAGGTGGCCCTTACGAGTGGATTCAGTGGCTTTTATCTCGGAATGG	594
DB	652	AAACAGAAGGATTTTCATAGAGATATTTATGGCTCAAACAATTTCTATTTCAAAGGCGAGTGG	711
QY	595	GGCCAGACCAAGCCCTATGCGGAATCCAAGGGGATTCAGATTAATCGGCGATATGCCCATC	654
DB	712	CAGCGAATTCGTAAATATGCAAAAGAGTTGGGTATCAGCAATAATGGTGCATGCCCNTA	771
QY	655	TTTGTGGCCTTCGATTTCTCAGATGTCTGGGCCAAACCCGACGTACTTCTACCTCGAGGCC	714
DB	772	TATGTTGGCTACCATAGTGCAGATGTTTGGCAACAAGGAATAATCGTTTTTGTCTGGACAAG	831
QY	715	GATGGCAACCCCAACGTTGGTGGCGGGGTTTCCGGGGACATCTTCTCCGAAACCGGCCAG	774
DB	832	AATGGTTTTCCGACTTTTGTAGTGGTGTTCACACCTGATGCAATTAAGTGAACCTGTCAG	891
QY	775	CTCTGGGGCAATCGCTCTATCGCTGGGATGTGATGGAAGGCAACAACTTTGCTCTGCTGC	834
DB	892	CTTTGGAAACAGTCCATTGTACGACTGGAAAGCTATGGAAAGCAGGTGGATTTGAATGGTGG	951
QY	835	ATTGCCCGCATAGGCAGTCCGCTCAAGCAGTGCACCTGGTGGCGCATCGACCACTTCCCGC	894
DB	952	ATAAAAAGGATTAATCGTGCCCTTGATTTGTATGATGAGTTCGATAGACCAATTTCCCGG	1011
QY	895	GGGTTTTGAAGCTACTGGGAGGTTCCGTTTTGGCCGCCCAATGCTGTGGAGGGCGCTGG	954
DB	1012	GGGCTTGCAGTTTTTGGGCAGTTTCCCTCTGAGTCAAAAGTAGCGCTGGTTGGAAGCTGG	1071
QY	955	GTCAAAAGCCCCAGGGGAGAAAGCTGTTTGCTTCGGTGGCGGCCCAACTGACGATCGCCCC	1014
DB	1072	AGGGCTGGACCGAGGAATGCAATTTTTTGATGCAATTAATCAAGCTGTCGGTAGAATAAC	1131
QY	1015	ATCATTTGCCAAGACCTGGGGGTGATCACCCCGAGGTGGAGGCTTTGGCCGATGGCTTC	1074
DB	1132	ATAATAGCAAGAACCTGGGGGTGATACAGAGATGTTGTTGACCTAAGGAAGTCTATT	1191
QY	1075	GGGTTCCCCGGCATGAAGATTTTTGCAGTTTGTCTTTTTCGGGTGAGGACAAACGCTTTTTG	1134
DB	1192	GAGCTCCTGGATGGCAGTTCTTCAGTTTGCTTTTGGCGGTGGTCTTGACAATCCACAC	1251
QY	1135	CCCCACAACATPACCCCGCGCACGGCAATGTGGTGGTGTACAGCGGAACCCACGACAGCAC	1194
DB	1252	TTGCCCATAAACCATG---AATTGTCAAGTAGTATACACCGGAACACACGCAACATGAC	1308
QY	1195	AC	1196
DB	1309	AC	1310

Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: xwing@genome.arizona.edu  
PCR Primers  
FORWARD: T7  
BACKWARD: M13r  
Plate: 0319 row: P column: 13  
Seq primer: M13r  
Class: BAC ends.

Location/Qualifiers  
1. 877  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="ZM5BB0319P13"  
/lab\_host="DH10B"  
/clone\_lib="ZM5BBB"  
/note="Vector: pBeloBAC11; Site 1: HindIII; Site 2: HindIII; Zea mays L. ssp. mays"

ORIGIN

Query Match 9.9%; Score 149; DB 10; Length 877;  
Best Local Similarity 52.5%; Pred. No. 4.9e-26;  
Matches 351; Conservative 0; Mismatches 315; Indels 3; Gaps 1;

Qy 571 CAGTGGCTTTTATCTGGAATGGGGCCAGACCAAGGCTATGCCAATCCAAGGGGATT 630  
|||  
Db 78 CAATTCTCTGTTCAAAGGCAGTGAAGAAATTCGGCACTATGCACAGCAAAAGGAGTC 137  
|||  
Qy 631 CAGATTATCGGCATATGCCATCTTTGTGGCCTTCGATTCTCTCAGATGTCGGGCCAAC 690  
|||  
Db 138 AGCTAATGGGACATGCTATTATGTAGGTTATCAGATGCTGATTTTGGGCAAAAT 197  
|||  
Qy 691 CCGCAGTACTTTCACCTCGAGGCGGATGGCAACCCACGGTGGTGGGGCGCTTCGCGG 750  
|||  
Db 198 AAGAAACAATTTTGTGTAATAGGCATGGTTATCTCTATGCTAGTCAAGTGTCTCTCCT 257  
|||  
Qy 751 GACTACTTTCGAAACCGGCAGCTCTGGGGCAATCCGCTCTATCGCTGGGATGTGATG 810  
|||  
Db 258 GATGCTTCAGTGATACTGGTCACTCTGGGGAGTCCTATATGATTGGAAGCCATG 317  
|||  
Qy 811 GAAAGGACAACTTTTCCTGGTGTCATTGCGCGATGAAGCAGTCGCTCAAGCAGTGCAC 870  
|||  
Db 318 GAGAAAGATGGATTTTATATGGTGATAGTCGATAGTACGCTGCACAGACTCTATGAT 377  
|||  
Qy 871 CTGCTGGCATCGACCACTTCCGCGGTTTGAAGCTACTGGAGGTTCCGTTTGGCCGG 930  
|||  
Db 378 GAATTCAGGATAGATCACTTTAGAGGATTTGCTGGCTTTTGGGCTGTCCTCTGAAGCG 437  
|||  
Qy 931 CCCAATGCTGTGAGGGGCGCTGGGTCAAGCCCCAGGGGAGNAGCTGTGTCGGGTG 990  
|||  
Db 438 AAAGTTGCAATGTTGGACGATGGAAGTGTGTCCTCAAAATCTTTTGTTCGACGCCATC 497  
|||  
Qy 991 CGGGCCCACTGAGCGATGCGCCCATCTTTCGGAAGACCTCGGGGTGATCACCCCCGAG 1050  
|||  
Db 498 TCCAAAGCAGTTGGAAGATTAATATCATAGCTGGAAGATCTGGAGTTATTACTGAGAT 557  
|||  
Qy 1051 GTGAGGCTTTGCGCATGGCTTCGGGTTCCTCCCGCATGAAGATTTTGCAGTTTGTCTTT 1110  
|||  
Db 558 GTAGTTCACTCAGGAATATATATGGAGCACCTGGAATGCTGTTCTCCAGTTTGGTTTT 617  
|||  
Qy 1111 TCCGGTGTAGGACAAACGCTTTTTCGCCCACTACCCCGGCGACGGCATGTGTGTG 1170  
|||  
Db 618 GGTAGTGACTCTAGCAACCCCTCAATTTACCTTCACAATCACGAGCA---CAATCAAGTTGTG 674  
|||  
Qy 1171 TAGAGGGAACCCACGACACGACACACCCTCGGATGTTCCGCAACCGCGCGGAGGCC 1230  
|||  
Db 675 TACACTGGACACATGACATGACATGACATGACATGACATGACATGACATGACATGACATG 734  
|||



397	AAAGTAGCACTGGTTGGAAGCTGGAGGCTGGACCAAGGAATAGCTTTTGGACAGCTC	456
991	CGGGCCCACTGACGATGCGCCCATCTGCGGAGACCTGGGGGTGATCACC	1050
457	TTCAAGCTGTTGTGAGAAATAGATATAATAGCAGAGATCTGGGGTAATTTACTGAAGAT	516
1051	GTGAGGCTTTGCGGATGCTTCGGGTTCGCCGATGAAGATTTTGCACTTTGCTTTT	1110
517	GTGCTTACGCTTAAGGAATATCCATTTGGGCGCCCTGGATGGCAGTTTCTCCAGTTTGCTTC	576
1111	TCCGCTGAGGACCAACGCTTTTGGCCCACTACCCCGGACCGCAATGTGTGGTG	1170
577	GGAGTGGTCTTGACCAACCTCATTTGGCCACACCAATGAAATGG---ATCAAGTTTG	633
1171	TACAGCGGAACCCACGACAGCACACACCTGGATGGTTCCGACCGCGCGGAGGCC	1230
634	TACACTGGGACACATGATAACGATACAGTTCCTTGGCTGGTGCGCAAAATTTACCAGAGGAG	693
1231	GA	1232
694	GA	695
RESULT 4		
CNS0A259	1875 bp mRNA linear HTC 06-FEB-2004	
LOCUS	Arabidopsis thaliana Full-length cDNA Complete sequence from clone	
DEFINITION	GSLTFB312F05 of Flowers and buds of strain col-0 of Arabidopsis	
thaliana (thale cress).		
ACCESSION	BX829726	GI:42459525
VERSION	1	1875
KEYWORDS	HTC; GSLT cDNA.	
SOURCE	Arabidopsis thaliana (thale cress)	
ORGANISM	Arabidopsis thaliana	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
AUTHORS	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;	
	rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	
TITLE	1 (bases 1 to 1875)	
	Castelli, V., Aury, J.M., Jaillon, O., Wincker, P., Clepet, C.,	
	Menard, M., Cruaud, C., Quetier, F., Scarpelli, C., Schachter, V.,	
	Temple, G., Caboche, M., Weissenbach, J. and Salanoubat, M.	
	Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:	
	A Combined Approach to Evaluate and Improve Arabidopsis Genome	
JOURNAL	Annotation	
REFERENCE	Unpublished	
AUTHORS	2 (bases 1 to 1875)	
TITLE	Genoscope.	
JOURNAL	Direct Submission	
COMMENT	Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :	
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr	
	- Web : www.genoscope.cns.fr)	
	The sequences are based on single pass reads.	
	Life Technologies (a division of Invitrogen) members carried out	
	full-length libraries construction : Temple G.	
	Genoscope members carried out sequencing and annotation : Castelli	
	V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,	
	Schachter V., Weissenbach J., Salanoubat M.	
	URGV INRA : Clepet C., Caboche M.	
	Annotation is based on the June 2003 version of the Arabidopsis	
	genome released by MIPS (Munich Information center for Protein	
	Sequences). 5 prime and 3 prime are assembled with Phrap.	
	http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EP/Full	
	length	
	http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.	
FEATURES		
source	Location/Qualifiers	
	1..1875	
	/organism="Arabidopsis thaliana"	
	/mol_type="mRNA"	
	/db_xref="taxon:3702"	
	/clone="GSLTFB312F05"	
	/tissue_type="Flowers and buds"	
	/ecotype="Col-0"	

/plasmid="pCMVSPORT_6"			
complement(1..1875)			
/gene="At5g64860"			
gene	Query Match	9.6%;	Score 145; DB 4; Length 1875;
ORIGIN	Best Local Similarity	50.6%;	Pred. No. 5.5e-25;
	Matches 403; Conservative	0;	Mismatches 390; Indels 4; Gaps 2;
QY	415	TGGCTGGAAGACTATGCGCTCTTTATGCGCCCTCAAGACCGCGTTTGAAGCGCAAGCCCTGG	474
Db	688	TGGCTTGAAGATGCTGCTTATTTTGCAGCTATTGACAATTTTAAATGATACAGTTGG	747
QY	475	AACGAGTGGAGCCCGAGCTGCGGACCGCTGAACCGGCTGCCCTGCCAGGCGCCGTGAG	534
Db	748	TTTGAGTGGCTGGAACCACTTAAAAACCGTCATCTTTCTGCTTGGACCTATATATGAA	807
QY	535	GAGCTGGCGGAGGAGGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTATCTGGAATGG	594
Db	808	AGCCAAAGGAGTTTATAGACTTGTTCATTGCTAAGCAATTTTGTTCAAAGGCAAGTG	867
QY	595	GCCCAAGCAAGGCTATGCGCAATCCAAAGGGATTCAGATTCGGCGATATGCCCATC	654
Db	868	CAGAAAGTTCTGTAGTATGACGCGGCGCAGGAGTTGATATATGCGGATATGCCCAIT	927
QY	655	TTTGTGGCTTCGATTCCTCAGATGCTGCGGCAACCCCGCAGTACTTCTACCTCGAGGCC	714
Db	928	TATGTAGGATATCAGATGCGACGCTTTGGGCAATTAAGAAACATTTCTTACTGNAACAG	987
QY	715	GATGGCAACCCCAACGCTGCTGCGGCGGCTTCGCGGAGTACTTCTCCGAAACCGGCGAG	774
Db	988	AAAGGCTTCTCTTCTTCTGTTAGCGGTGTTCTCTGACTTGTTCAGTGAACACTGGTCAA	1047
QY	775	CTCTGGGCAATCCGCTCTATCTGCTGGATGTGATGGAAGGACAACTTTGGCTGTGTGC	834
Db	1048	CTGTGGGCGAGCCCTCTTTATGACTGGAAGCAATGGAGATGACCAATATCTTGTGTGG	1107
QY	835	ATTGCCCGCATAGGCAAGCTGCTCAAGCAGTGCACCTGGTGGCGCATCGACCACTTCCGC	894
Db	1108	GTTAATCGAATAAGACGCGCACAGGACTTGTATGACGAATGCAGGATTCATCTTCAGA	1167
QY	895	GGGTTTGAAGCCCTACTGGAGGTTCCGTTTGGCGGCCCAATCTGTGAGAGGGCGGTGG	954
Db	1168	GGATTTGAGGGTTTGGGCGGTCCCTCTGAAGCGAAGTTGCCATGTTGTCACCATGG	1227
QY	955	GTCAAAAGCCCGAGGAGAGCTGTTTGTGCGGTGCGGCGCCCACTGACGAGTGCSCCC	1014
Db	1228	AAGTAGGACCTGGAAGTCAATTTTATGATGCCATTTTAAAGGCGTTGGGAAGATCAA	1287
QY	1015	ATCATTTGCCGAGACCTGGGGGTGATCACCCCGAGGTGGAGGCTTTTGGCGGATGGCTT	1073
Db	1288	AATCATAGCTGAAGATTTGGGAGTTTAAAGATGTAGTTGAGCTGAGGAATCTAT	1347
QY	1074	CGGTTTCCCGCGCATGAAGATTTTGCACTTTTTCGGTGTAGGACAAACCGCTTTT	1133
Db	1348	CGAGCACTTGGATGGCCCTCTCCAAATTTGCTTTTGGAGGAGGCGCCGATACCCACA	1407
QY	1134	GCCCCCAACTACCCCGCGCACCGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1193
Db	1408	TTTACCTCACAATCATG---AAGTAACCAAGTTGTTTCTCTGTGTGTGTGTGTGTGTGTGT	1464
QY	1194	CACACCTGGATGGT	1210
Db	1465	CACATTTGAGGCTGGT	1481
RESULT 5			
LOCUS	CV470446	762 bp	mRNA linear EST 01-OCT-2004
DEFINITION	43319.1 Common Scab-Challenged Tubers Solanum tuberosum cDNA clone		
	43319 5', mRNA sequence.		
ACCESSION	CV470446		
VERSION	CV470446.1	GI:53695222	

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. .762
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Shepody"
/db_xref="taxon:4113"
/clone="43319"
/tissue_type="Tubers"
/lab_host="XL10-Gold"
/clone_lib="Common Scab-Challenged Tubers"
/notes="Vector: pBluescript II SK(+) XR; Site 1: EcoRI; Site 2: XhoI; supplier: Pathogen-challenge series. Tubers from pathogen-free Solanum tuberosum var. Shepody, clone 1756, were inoculated with Streptomyces scabiei EF-35, by applying 1 ml of a spore stock (OD650 = 0.7). Tubers were placed in trays of sand and watered from the bottom for the remainder of the experiment. Inoculated tubers were sampled at 2 days, 1 week, 2 weeks and 4 weeks post-infection. Harvested tubers were washed in distilled water, blotted dry, cut into small pieces and frozen in liquid nitrogen. RNA was isolated from each sample collection, pooled and used for cDNA library construction."
ORIGIN
Query Match 9.5%; Score 143; DB 7; Length 762;
Best Local Similarity 52.0%; Pred. No. 1.5e-24;
Matches 345; Conservative 0; Mismatches 315; Indels 3; Gaps 1;
QY 571 CAGTGGCTTTTATCTGGATGGGCGACCAAGCCCTATCCGAATCCAAAGGGATT 630
DB 98 CAGTCTTGTTCACGCAATGGAAAAAGTTCTGTACTATGCAGATCCAAAGGAATC 157
QY 631 CAGATTATCGGCGATATGCCATCTTTGTGGCCTTCGATTCCTCAGATGTCGGCCAAC 690
DB 158 AGTATTAATGGAGACATGCCAATATGTTGGATATCAGATGCTGATGTTTGGGCCAAC 217
QY 691 CGCAGTACTTCTACCTCGAGGCGGATGGCAACCCACCGGTGGTGGCGGCGTTCGCGG 750
DB 218 AAGAACAAATTTTGTGTAATAGGAAGTTTCCCTCTTATAGTTAGTGTCTCTCA 277
QY 751 GACTACTTCTCGAAACCGGCCAGCTCTCGGGCAATCCGCTCTATCGTGGGATGATG 810
DB 278 GACGCTTTAGTGAACCTGTCACATATGCGGCGAGCCCTCTCTATGATTGGAAGCCATG 337
QY 811 GAAGGGACAACTTTCCTGGTGCATTCGCCGATAGGCGATGCGTCAAGCAGTCCAC 870
DB 338 GAGAGGATGGATTTTTCATGGTGGGTATGCCGAATTCACAGTGCACACGATCTTTTGTAT 397
QY 871 CTGGTGGCATCGACCACTTCCCGGGTTTGAAGCCTACTTGGAGGTTTCGTTTGGCCG 930
DB 398 GAATTTAGATAGATCATTAGAGANTTTGCGAATTTTGGCTGTTCTTCTGAGGAA 457
QY 931 CCCAATGCTGTGAGGGCGCTGGGTCAAGCCCGAGGAGAAAGCTGTTTGTGCGGTG 990
DB 458 AAAATTGCAATTCGCGGACGGTGGAAAGGTGGGACCTGGAAAAACCTTTGTTGATGCTATC 517

```

EST.  
Solanum tuberosum (potato)  
Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Solanales; Solanales; Solanaceae; Solanum.  
1. (bases 1 to 762)  
Flinn, B., Rothwell, C., Sardana, R., Griffiths, R., Laque, M., De Koeber, D., Audy, P., Goyer, C., Li, X.-O., Wang-Pruski, G. and Regan, S.  
Generation of ESTs from common scab-challenged potato tubers  
Unpublished (2004)  
Contact: Barry Flinn  
The Canadian Potato Genome Project - BioAtlantech  
921 College Hill Rd, Fredericton, ON, B3B 6Z9, CANADA  
Email: bflinn@bioatlantech.nb.ca  
Seq primer: T3.

Location/Qualifiers  
1. .762  
/organism="Solanum tuberosum"  
/mol\_type="mRNA"  
/cultivar="Shepody"  
/db\_xref="taxon:4113"  
/clone="43319"  
/tissue\_type="Tubers"  
/lab\_host="XL10-Gold"  
/clone\_lib="Common Scab-Challenged Tubers"  
/notes="Vector: pBluescript II SK(+) XR; Site 1: EcoRI; Site 2: XhoI; supplier: Pathogen-challenge series. Tubers from pathogen-free Solanum tuberosum var. Shepody, clone 1756, were inoculated with Streptomyces scabiei EF-35, by applying 1 ml of a spore stock (OD650 = 0.7). Tubers were placed in trays of sand and watered from the bottom for the remainder of the experiment. Inoculated tubers were sampled at 2 days, 1 week, 2 weeks and 4 weeks post-infection. Harvested tubers were washed in distilled water, blotted dry, cut into small pieces and frozen in liquid nitrogen. RNA was isolated from each sample collection, pooled and used for cDNA library construction."

ORIGIN  
Query Match 9.5%; Score 143; DB 7; Length 762;  
Best Local Similarity 52.0%; Pred. No. 1.5e-24;  
Matches 345; Conservative 0; Mismatches 315; Indels 3; Gaps 1;  
QY 571 CAGTGGCTTTTATCTGGATGGGCGACCAAGCCCTATCCGAATCCAAAGGGATT 630  
DB 98 CAGTCTTGTTCACGCAATGGAAAAAGTTCTGTACTATGCAGATCCAAAGGAATC 157  
QY 631 CAGATTATCGGCGATATGCCATCTTTGTGGCCTTCGATTCCTCAGATGTCGGCCAAC 690  
DB 158 AGTATTAATGGAGACATGCCAATATGTTGGATATCAGATGCTGATGTTTGGGCCAAC 217  
QY 691 CGCAGTACTTCTACCTCGAGGCGGATGGCAACCCACCGGTGGTGGCGGCGTTCGCGG 750  
DB 218 AAGAACAAATTTTGTGTAATAGGAAGTTTCCCTCTTATAGTTAGTGTCTCTCA 277  
QY 751 GACTACTTCTCGAAACCGGCCAGCTCTCGGGCAATCCGCTCTATCGTGGGATGATG 810  
DB 278 GACGCTTTAGTGAACCTGTCACATATGCGGCGAGCCCTCTCTATGATTGGAAGCCATG 337  
QY 811 GAAGGGACAACTTTCCTGGTGCATTCGCCGATAGGCGATGCGTCAAGCAGTCCAC 870  
DB 338 GAGAGGATGGATTTTTCATGGTGGGTATGCCGAATTCACAGTGCACACGATCTTTTGTAT 397  
QY 871 CTGGTGGCATCGACCACTTCCCGGGTTTGAAGCCTACTTGGAGGTTTCGTTTGGCCG 930  
DB 398 GAATTTAGATAGATCATTAGAGANTTTGCGAATTTTGGCTGTTCTTCTGAGGAA 457  
QY 931 CCCAATGCTGTGAGGGCGCTGGGTCAAGCCCGAGGAGAAAGCTGTTTGTGCGGTG 990  
DB 458 AAAATTGCAATTCGCGGACGGTGGAAAGGTGGGACCTGGAAAAACCTTTGTTGATGCTATC 517

QY 991 CGGGCCCAACTGAGCGATCGGCCCATCATTTGCCAGAGACCTGGGGGTGATCACCCCGAG 1050  
DB 518 TTACAAGCTGTTGGGAAGATCAATATATAGCAGAGACTTGGAGTAATATCCGAGAC 577  
QY 1051 GTGAGGCTTTGGCGCATGGCTTCGGGTTCCTCCGGCATGAAGATTTTTCAGATTTGCTTTT 1110  
DB 578 GTTGTTCAGCTAAGAAAGTCCATTGAGGCACCTCGAATGCTGTACTCCAGTTTGCATTT 637  
QY 1111 TCCGCTGAGGACAAACCCCTTTTGGCCCCCACTATCCCGCGCAGCGCAATGTTGTTG 1170  
DB 638 GGCACTGAGCGCTGAACACCCCTCATTTACCTCACAATCATGAGCAGAACCAAGTA---GTG 694  
QY 1171 TACAGCGGAACCCACCAACAGCACACACCCCTGGATGTTCCGACCGCGCCGAGGCC 1230  
DB 695 TATACCTGGACACATGACATGATGATGATCGAGGTGGTGGGATCTTTTGGCCACAGAA 754  
QY 1231 GAG 1233  
DB 755 GAG 757

RESULT 6  
AJ795827  
LOCUS  
DEFINITION  
AJ795827 733 bp mRNA linear EST 08-DEC-2004  
018.3.09.d02, mRNA sequence.  
ACCESSION  
AJ795827  
VERSION  
AJ795827.1 GI:51111155  
SOURCE  
EST.  
ORGANISM  
Antirrhinum majus (snapdragon)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Lamiales; Plantaginaceae; Antirrhineae; Antirrhinum.  
REFERENCE  
1 (bases 1 to 733)  
Bey, M., Stueber, K., Fellenberg, K., Schwarz-Sommer, Z., Sommer, H., Saedler, H. and Zachgo, S.  
Characterization of Antirrhinum Petal Development and Identification of Target Genes of the Class B MADS Box Gene DEFICIENS  
Plant Cell 16 (12), 3197-3215 (2004)  
15539471  
COMMENT  
Contact: Schwarz-Sommer Z  
Molekulare Pflanzen-genetik  
MPI fuer Zuechtungs-forschung  
Carl-von-Linne Weg 10, D-50829, Germany.  
Location/Qualifiers  
1. .733  
/organism="Antirrhinum majus"  
/mol\_type="mRNA"  
/db\_xref="taxon:4151"  
/clone="018.3.09.d02"  
/tissue\_type="whole plant"  
/clone\_lib="Antirrhinum majus whole plant"

ORIGIN  
Query Match 9.3%; Score 140.2; DB 1; Length 733;  
Best Local Similarity 54.6%; Pred. No. 7.4e-24;  
Matches 280; Conservative 0; Mismatches 233; Indels 0; Gaps 0;  
QY 608 CCTATGCCAATCCAAAGGGATTCAGATTCATCGCGATATGCCATCTTTGTGGCCTTCG 667  
DB 207 CTTATGCTCAGATGAGAGAAATCAGTATTAATGGAGATATGCTATATATGTTGGTTATC 266  
QY 668 ATTCTCAGATGCTGGGCGCAACCCGCGACTACTTCTACCTCGAGGCGGATGGCAACCCCA 727  
DB 267 ATAGTGCAGATGCTGGGCTATNAGAACATTTTGTCTTAATAGCAGTGGCTTTCCCC 326  
QY 728 CGGTGGTGGCGGGCGTTTCGCGGAGTACTTCTCCGAAACCGGCGCAGCTCTCGGGGCAATC 787  
DB 327 TTCTAGTTAGTGGTGTTCCTCCCTGATGCTTCAGCGAGACTGTCAGCTTTTGGAAACAGCC 386



```

method 4 described in Bonaldo et al., (1996) Genome
Research 6: 791-806."

ORIGIN
Query Match          9.2%; Score 138.2; DB 3; Length 538;
Best Local Similarity 56.4%; Pred. No. 2.2e-23;
Matches 254; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

QY 468 GGCCTGGAAACAGTGGAGGCCCGAGTGGCGACCGTGAACCGGCTGCCCTGGCCAGGCG 527
DB 82 GCGGTGGTGGCAGTGGCGGAGCCGTTGCGGTGGCGGAGAGGAGGCGCTCAAGGAGTT 141
QY 528 CCGTGGAGAGTGGCGGAGGAGTGGCGCTTACAGAGTGAATTCAGTGGCTTTTATCT 587
DB 142 CCGGGAGAGCAAAAGGAGCGCAATTGACAGTTCGTTGTGATCCAGTACTTCTTCGAGAA 201
QY 588 GGAATGGGGCCAGACCAAGGCCCTATGCCGAATCCAAAGGGGATTCAGATATTCGGGATAT 647
DB 202 GCAGTGGAAAGCGATCCGCTCTACCGGAAACGCGAGGCGCATCAAACTCATCGCGCAT 261
QY 648 GCCCATCTTTTGGGCTTCGATTCCTCAGATGTCTGGGCAACCCCGAGTACTTCTACCT 707
DB 262 GCCCATCTACGTGGGCGGCACAGCGCAGATGTGGGCGCAACCCCGACCTGTTCGAGCT 321
QY 708 CGAGGCCGATGGCAACCCAGCGTGGTGGGCGGCTTCCGGGAGTACTTCTCCGAAAC 767
DB 322 GAACGAGGCGGCGCTGCGCGGAGCGAGTCCGCGGAGTCCCGGAGCGCTTCTCAGCAAC 381
QY 768 CGGCCAGCTCTGGGCAATCCGCTCTATCGTGGGATGTGATGGAAGGCAACATTTTC 827
DB 382 GGGCCAGCTGTGGGCGGCGGCTTACAGTGGCGGCCCAACAGAGAGGAGGCTTCAA 441
QY 828 CTGGTGCAATGGCCCGCAATAGGAGTGGTCTCAAGAGTGGCCAGTCTGGTGGCATCGACCA 887
DB 442 GTGGTGAGCGGCGCATGGCCGAGCTTGGGAGCTGTACGAGAGTGCNNATCGACCA 501
QY 888 CTTCGCGGGTTTGAAGCTTACTGGGAGGT 917
DB 502 CTTCGCTGGCTTCGAGGCTTACTGGAGCGT 531

RESULT 8
AY109140 Zea mays PC0094517 mRNA sequence.
LOCUS Zea mays 1155 bp mRNA linear HTC 25-FEB-2005
DEFINITION Zea mays PC0094517 mRNA sequence.
ACCESSION AY109140
VERSION AY109140.1 GI:21212571
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1155)
Gardner, J., Schroeder, S., Polacco, M.L., Sanchez-Villeda, H.,
Fang, Z., Morgante, M., Landewe, T., Fengler, K., Uscache, F.,
Hanafey, M., Tingey, S., Chou, H., Wing, R., Soderlund, C. and Coe, E.H.
Jr.
Anchoring 9,371 maize expressed sequence tagged unigenes to the
bacterial artificial chromosome contig map by two-dimensional
overgo hybridization
Plant Physiol. 134 (4), 1317-1326 (2004)
15020742
2 (bases 1 to 1155)
Hainey, C.P., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
3 (bases 1 to 1155)
Coe, E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of

QY 788 CGCTCTATCGTGGGATGTGATGAAAGGCAAACTTTGCTGTGATTCGCGCATAA 847
DB 387 CTCGTATGATGAAAGCAATGAAATAGATGATTTTCATGTTGATTAAGCGTATAA 446
QY 848 GGCAGTCTGCTCAAGCAGTGCACCTGGTGGCGATCGACCACTTCCCGGGTTTGAAGCCT 907
DB 447 GAGCTGCACAGATCTATTTCGATGAATTCGGGATCGACCACTTCAGAGGATTCGCTGGCT 506
QY 908 ACTGGAGAGTTCGTTTGGCGGCCCAATCTCTGTGGAGGGCGCTGGTCAAGAGCCAG 967
DB 507 TTTGGGCTGTCTCTTCTGAAGCAAAAGTTCGGATGGTTGGGAAATGAAAGGAGGAGCTG 566
QY 968 GGCAGAGAGCTGTTTGTCTGGCGTGGCGGCCCACTGAGCGATGCGCCCATCATTTGCCAAG 1027
DB 567 GAGATCTCTATTGTAGTCTATCTTTAGAGCTGTTGGAGATATCAATATCGTAGCAGAG 626
QY 1028 ACCTGGGGGTGATCACCCCGAGTGGAGGCTTTTGGCGGATGGCTTCGGGTTCCCGGCA 1087
DB 627 ACTTGGGGTCACTACTGAAGATGTAGTGCAGCTTAGGAGGTCCATCGGAGCACCTGGAA 686
QY 1088 TGAAGATTTGCAAGTTCGTTTTCGCGTGAGG 1120
DB 687 TGGCGGCTCCAGTTGGCTTCGAGAGTGATG 719

RESULT 7
BI720945 538 bp mRNA linear EST-19-SEP-2001
LOCUS 1031053A09.y1 C. reinhardtii CC-1690, Stress II (normalized),
DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BI720945
VERSION BI720945.1 GI:15696640
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
1 (bases 1 to 538)
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C.,
Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1031
Unpublished (2001)
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
Location/Qualifiers
1. 538
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress II
(normalized), Lambda Zap II"
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; Stress condition II library, constructed by John
Davies and Jeffrey McDermott, combines cDNAs from CC-1690
cells grown to mid-log phase in TAP (NH4+ - containing)
and shifted to TAP - NO3- (24hrs); H2 production
conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
Zap II (Stratagene) in the EcoRI (5') and XhoI (3')
sites. pBluescript II SK- plasmids were excised from the
lambda Zap clones by superinfection with Exsist
(Stratagene) phage. The library was normalized using

```

Missouri, Columbia, MO 65211, USA  
 If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES  
 source  
 1. 1155  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /db\_xref="MaizeGDB:635393"  
 /db\_xref="taxon:4577"  
 /clone\_lib="Maize Mapping Project/DuPont Corneaus Library"  
 /notes="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

## ORIGIN

Query Match 8.8%; Score 132.4; DB 4; Length 1155;  
 Best Local Similarity 52.2%; Pred. No. 7.1e-22;  
 Matches 318; Conservative 0; Mismatches 288; Indels 3; Gaps 1;

QY 624 GGGATTTCAGATTATCGCGGATATGCCCATCTTTTGTGGCTTCGATTCCTCAGATGCTG 693  
 DB 53 GTGTATCAGCATCATGGTGCATGCTGTATATGTGGCTACCATAGCGCATATGTTG 112  
 QY 684 GGCACACCCGACATCTTACCTCGAGCGCGATGCGCAACCCACCGTGTGCGGCGGT 743  
 DB 113 GCGCAACAGGAATCATTTTTTGTGCAAAAACCGTTTCCCACTTTTGTAGTGGCGT 172  
 QY 744 TCCGCGGACTACTTCTCCAAACCGCCAGCTCTGGGCAATCCGCTCTATCGCTGGGA 803  
 DB 173 TCACCTGATGCAATTTAGTGAACGGGTCACTATGGAACAGTCCATTTGACACTGGAA 232  
 QY 804 TGTGATGGAAGGAGCAACTTTCCCTGGTGCATTTGCCCGCATAGGCAAGCTCGCTCAAGCA 863  
 DB 233 AGCTATGGAAGCAGATGTTTTCATGTTGATATAAGAGGATTAACGCTGCTTCATTT 292  
 QY 864 GTCCACCTGGTGGCATCGACACTTCGCGGGTTTGAAGCCATCTAGGAGTTCGGTT 923  
 DB 293 GTATGATGAATTCGGTATTTGACCAATTTCCGCGGGCTTGGCGGTTTGGGCACTCCCTTC 352  
 QY 924 TGGCGCGGCCAATGCTGTGGAGGGCGCTGGGTCAAAAGCCCGAGGAGAGCTGTTTGC 983  
 DB 353 TGATGCAAAAGTAGCACTGTTGGAAGCTGGAGGGCTCGNCCAGGAATAGCTTTTGA 412  
 QY 984 TCGGTGCGGGGCCAACTAGCGCATGCGCCCATCATTTGCGGAAGACTGGGGTGATCAC 1043  
 DB 413 CAGCTCTTCAAAGCTGTTGGTAGATAGATATATAATAGCAGAGATCTGGGGTAATTAC 472  
 QY 1044 CCCCAGGTGGAGGCTTTGGCGCATGCGTTCGGGTTCCCGCATGAAGATTTGCAATT 1103  
 DB 473 TGAAGATGTCGTTACGTAAGGAATTCATTTGGGGCCCTGGGATGGCAGTTCTCCAGTT 532  
 QY 1104 TGTCTTTTCCGGTGAGGACAAAGCCCTTTTGGCCCACTACTCCCGCAGCGGAATGT 1163  
 DB 533 TGTCTTCGAGGTGGTCTGCAACCCCTCATTTGGCCACACCAATG---AATGATCA 589  
 QY 1164 GGTGTGTACAGCGGACCCACGACCAACGACACACCCCTGGATGTTCCGCAACCGGCC 1223  
 DB 590 AGTTGTGTACACTGGAAACACATGATACGATACAGTCTTCTGGCTGTGGCAAAATTACC 649  
 QY 1224 GGAGGCCGA 1232  
 DB 650 AGAGGAAGA 658

## RESULT 9

CK767135  
 LOCUS  
 DEFINITION

Persea americana (avocado)

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana

Persea americana



Clontech's PCR Suppression Subtractive Hybridization was used. Tester = plants sprayed with 50mM methyl jasmonate, and tissue harvested 1,8, 24 and 48h after treatment; driver = plants sprayed with water and tissue harvested 1,8, 24 and 48h. Cloned into Advantage PCR cloning vector (Clontech)."

ORIGIN

Query Match 8.4%; Score 126.8; DB 6; Length 683;  
Best Local Similarity 54.1%; Pred. No. 1.6e-20;  
Matches 331; Conservative 0; Mismatches 267; Indels 14; Gaps 3;

QY 610 TATGCGAATCAAGGGGATTCAGATTATCGCGGATATCGCCATCTTTGTGGCTTCGAT 669  
DB 71 TATGACGGCGGAGAGATTGATATATGGAGATATGCCCATTTATATAGGATATCAC 130  
QY 670 TCCTCAGATGTCTGGGCAACCCGCGAGTACTTCTACCTCGAGCCGATGGCAACCCACG 729  
DB 131 AGTGCAGAGCTTTGGGCAATAAAGAAACATTTCTTACTGAACAAGAAAGCTTTCTCTT 190  
QY 730 GTGGTGGCGGGCTTCCGCGGAGTACTTCTCGAAGCCGCGGAGCTCTGGGGCAATCG 789  
DB 191 CTGTGTAGCGGTCTCTCTCTGCTGACTTTGTTTCACTGGAACCTGGTCACTGTGGGGCAGCCCT 250  
QY 790 CTCTATCGTGGGATGTGATGGAAGGGAACAATTTGCTGTGTGCTATTTGCCGCAATAAG 849  
DB 251 CTTTATGACTGGAAGCAATGGAGAGTGACCAATATTTCTTGTGGGTAAATCGAATAAGA 310  
QY 850 CAGTCGCTCAAGCAGTCCACCTGTGTGCGCATCGACCACTTCCGCGGGTTTGAAGCCTAC 909  
DB 311 CGGCAAGAGCTTTGATGCGAATTCAGGATTTGATCAAGAGATTTGCAGGGTTT 370  
QY 910 TGGAGGTTCGGTTTGGCGGCGCAATGTGTGAGGGGGCGTGGGTCAAGCCCGAGGG 969  
DB 371 TGGCGGTCTCTCTCTGAGCGAAGTTGCCATGTTGGAGAGTGG---AAGGACCTGGA 426  
QY 970 GAGAAGCTGTGTGCTGCGGTGGGGCCCAACTGAGCGATGCGGCCCATCTTTCGCGAAGAC 1029  
DB 427 AGTCAATATTTGATGCCATTTCCAAAGGGCTTGGGAAGATCAAAATCATAGCTGAAGAT 486  
QY 1030 CTGGGCGTATCAACCCCGAGGTGGAGGCTTTCGCGCATGGCTTCGGGTTCCCGCGCATG 1089  
DB 487 TTGGAGTTATTAATAAGATGTAGTTAGCTGAGGAAATCTATCGAGCAGCTTGGAAATG 546  
QY 1090 AAGATTTGCAATTTGCTTTTCCGCTGAGGAC----AAGCCTTTTGGCCCCCAACTA 1145  
DB 547 GCGCTCTCCAAATTTGCTTTTGGAGGAGGCGCGCGATTAACCCACATTTACTCACAATCA 606  
QY 1146 CCGCGCGCAGGCAATGTGGTGTGTACAGCGGAACCCAGCAACACCAACCCCTGGG 1205  
DB 607 -----TGAAGTAAACCAAGTTGTATCTCTGGAATCATGACACACATATTCGAG 660  
QY 1206 ATGTTCCGCGAC 1217  
DB 661 CTGGTGGGACAC 672

RESULT 11

COL19845 646 bp mRNA linear EST 16-JUN-2004  
LOCUS GR\_EB022P17.r GR\_Eb Gossypium raimondii cDNA clone GR\_EB022P17  
DEFINITION 3', mRNA sequence.  
ACCESSION COL19845  
VERSION COL19845.1 GI:48818532  
KEYWORDS EST.  
SOURCE Gossypium raimondii  
ORGANISM Gossypium raimondii  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.  
1 (bases 1 to 646)  
Kim,H., Yu,Y., Kudrna,D., Hatfield,J., Stum,D., Mueller,C.,  
Udall,J.A., Rapp,R.A., Wendel,J.F., Rao,K., Soderlund,C. and

DB 241 GAATTTCCGATAGACCAATTTTCGAGGACTTGTGGCTTTTGGCTGTCTCTGAATCT 300  
QY 931 CCAATGCTGTGGAGGGCTGGTCAAGCCCGAGGAGAGCTTTTGTGCGGGTG 990  
DB 301 AAGATTGCTATGTTCCGAAGATGAAGCTGGACCAACAAAGCATTTTGTATGCCGTT 360  
QY 991 CGGGCCCACTGAGCGATGCGCCCATCTATTTCCGCAAGACCTGGGGGTGATCACCCCGAG 1050  
DB 361 TTCAAGCTGTTGGAAGATCAATATTTGCTGAGGACTTGGAGTCAATCACAGAGGAC 420  
QY 1051 GTGAGGCTTTGGCGGATGCTGGCTTCCCGGCGATGAAGATTTTGCAGTTGCTTTT 1110  
DB 421 GTAATTGAGTAAAGAAAGCTATTTGTCACCAAGGATGGCAGTTCTCCAATTTGCTTTT 480  
QY 1111 TCCGTTGAGGACAGGCTTTTGGCCCAACAATCTACCCCGCAGCAGCAATGTGGTGTG 1170  
DB 481 GGGAGTGATTTCTGCAATCTCTATTTACCTCATATCATGAATG---ATCAAGTCGTA 537  
QY 1171 TACAGCGGAACCCACGACCAACGACCAACCTCTGGGATGTTCCGCAACCCGCGGAGGCC 1230  
DB 538 TACACTGGAACACATGATATGACACTGTGCTCGTGTGGTGGAGGAACTTGCAAGAGAA 597  
QY 1231 GAGCGGACC 1239  
DB 598 GAGAAGTCC 606

RESULT 10

CB074139  
LOCUS CB074139.1 GI:32362520  
DEFINITION EST.  
Arabisopsis thaliana (thale cress)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (bases 1 to 683)  
Mahalingam,R., Gomez-Buitrago,A.M., Eckardt,N., Shah,N.,  
Guevara-Garcia,A., Day,P.M., Raina,R. and Fedoroff,N.V.  
Characterizing the stress/defense transcriptome of Arabidopsis  
Genome Biol. (2003) In press  
Contact: Nina V. Fedoroff  
Biotechnology Institute  
Penn State University  
519 Wartik Lab, University Park, PA 16802, USA  
Tel: 8148635717  
Fax: 8148631357  
Email: nvf@psu.edu  
Sequence does not include adaptor sequences (corresponding to  
Clontech PCR-Select Adaptors 1 and 2R) present on either side of  
insert.  
Insert Length: 683 Std Error: 0.00  
Seq primer: Clontech PCR-Select Nested Primer 2R  
High quality sequence start: 15  
High quality sequence stop: 675  
POLYA=No.

FEATURES

Location/Qualifiers  
1..683  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/ecotype="Columbia"  
/db\_xref="taxon:3702"  
/clone="MJF-FB10"  
/tissue\_type="Leaf"  
/clone\_lib="Methyl Jasmonate Treated Arabidopsis  
Forward-Subtracted Library"  
/note="Vector: Advantage PCR cloning vector (Clontech);

Wing, R.A.  
Global assembly of Cotton ESTs  
Unpublished (2004)  
Contact: Rod A. Wing  
Arizona Genomics Institute  
The University of Arizona  
Forbes Building Room 303, Tucson, AZ, 85721-0036, USA  
Tel: 520 626 9595  
Fax: 520 621 1259  
Email: rwing@genome.arizona.edu  
Plate: 022 row: P column: 17.

TITLE Global assembly of Cotton ESTs  
JOURNAL Unpublished (2004)  
COMMENT Contact: Rod A. Wing  
Arizona Genomics Institute  
The University of Arizona  
Forbes Building Room 303, Tucson, AZ, 85721-0036, USA  
Tel: 520 626 9595  
Fax: 520 621 1259  
Email: rwing@genome.arizona.edu  
Plate: 022 row: P column: 17.

FEATURES  
source  
1. 646  
Location/Qualifiers  
/organism="Gossypium raimondii"  
/mol\_type="mRNA"  
/db\_xref="taxon:29730"  
/clone="GR\_EB022P17"  
/tissue\_type="floral"  
/dev\_stage="3 to +3 DPA"  
/lab\_host="DH10B"  
/clone\_lib="GR\_Eb"  
/note="Vector: pCMV SPORT-6.1; Site 1: NotI; Site 2: EcoRV; Library made by invitrogen with RNA supplied by Wendle lab. Directional cloned into NotI-EV. Colonies plated/picked by AGI. More glycerol clones held in -80."

## ORIGIN

Query Match 8.3%; Score 125.4; DB 7; Length 646;  
Best Local Similarity 51.9%; Pred. No. 3.6e-20;  
Matches 282; Conservative 0; Mismatches 261; Indels 0; Gaps 0;

QY 571 CAGTGGCTTTTATCTGGATGGGGCCAGACCAAGCGCTATGCGAATCAAGGGGATT 630  
DB 66 CAATCTCTTCCAAAGCAGTGAAGAAATTCGCACTATGACACAGAAAGGAGTC 125  
QY 631 CAGATTATCGCGGATATGCCATCTTTGTGGCTTCGATTCCTCAGATGTCTGGCCAAAC 690  
DB 126 AGCATATGGAGACATGCTTATTTATGTAGTTATCACAGTGTGTATTTGGGCAAT 185  
QY 691 CCGCAGTACTTCTACCTCGAGCGCGATGGCAACCCAGCGTGTGGCGGGTTCGCCGG 750  
DB 186 AAGAAACATTTTGTCTGAATAGGCATGGTTATCTATGCTAGTCACTGGTGTCTCTCT 245  
QY 751 GACTACTTCTCGAAACCGCCAGCTCTGGGGCAATCGCTCTATCGCTGGGATGTGATG 810  
DB 246 GATGCTCTAGTATCTGTGCTGCTGGGGAGTCTCTATGATTTGAAAGCCATG 305  
QY 811 GAAAGGACAACTTTGCTGCTGTCATTCGCCGATAGGCGAGTCTCAAGCAGTGGCCAC 870  
DB 306 GAGAAAGATGATTTTATGTGGATAAGTCGCATTAGACGTGCACAGACTCTATATGAT 365  
QY 871 CTGCTGGCATCGACACTCTCCGGGGTTTGAAGCCTACTGGGAGGTTCGTTTGGCCGG 930  
DB 366 GAATTCAGGATAGATCACTTTAGAGGATTTGCTGGCTTTTGGGCTGTCCCTTCTGAAGCG 425  
QY 931 CCCAATGCTGTGGGGGCTGGTCAAGCCCGAGGAGAGCTGTTTGTGGGTG 990  
DB 426 AAGTTGCAATGTTGGACCATGAAGTTGGTCTCTCAAAATCTTTGTTTCGACGCCATC 485  
QY 991 CGGGCCCACTGAGCGATGCGCCCATCATTTCCGGAAGACCTGGGGGTGATCACCCCGAG 1050  
DB 486 TCCAAAGCAGTTGGAAGATTAATATCATAGCTGAAGATCTGGGAGTTAATCTGAGGAT 545  
QY 1051 GTGAGGCTTTGGCGGATGCTGGGTTCGCCGCGATGAAGATTTTTCAGTTTGTCTTTT 1110  
DB 546 GTAGTTTCAGCTCAGGAAATATATGGAGCACCTGGATGGCTGTCTTCCAGTTTGTGTTG 605  
QY 1111 TCC 1113  
DB 606 TCC 608

RESULT 12

BE515486/c  
LOCUS  
DEFINITION  
WHE0613\_B06\_C112A Wheat ABA-treated embryo cDNA library Triticum aestivum cDNA clone WHE0613\_B06\_C11, mRNA sequence.  
ACCESSION  
BE515486  
VERSION  
BE515486.1 GI:9739432  
KEYWORDS  
EST.  
SOURCE  
ORGANISM  
Triticum aestivum (bread wheat)  
Triticum aestivum

REFERENCE  
AUTHORS  
Anderson, O.D., Chao, S., Han, P.S., Hsia, C.C., Johnson, R.R., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L., Tong, J.C., Verhey, S.D. and Walker-Simmons, M.K.  
TITLE  
The structure and function of the expressed portion of the wheat genomes - ABA-treated embryo library  
JOURNAL  
Unpublished (2000)  
COMMENT  
Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20  
Seq primer: Clontech Matchmaker 3' AD primer.  
Location/Qualifiers  
1. 605  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Brevor (soft, white, winter, common wheat)"  
/db\_xref="taxon:4565"  
/clone="WHE0613\_B06\_C11"  
/tissue\_type="Seed embryo"  
/dev\_stage="Mature dormant seeds"  
/lab\_host="E. coli DH12S"  
/clone\_lib="Wheat ABA-treated embryo cDNA library"  
/notes="Vector: pGAD10; Site 1: EcoRI; Site 2: XhoI; Embryos were cut from mature, dormant seeds and imbibed in 25 micromolar ABA (abscisic acid) in 5 mM Mes buffer, pH 5.7, for 12 hr at 22 C. The tissue, total RNA, and poly(A) RNA were prepared by Steven Verhey in M.K. Walker-Simmons's lab (USDA-ARS, Washington State Univ., Pullman, Washington 99164-6420. A cDNA library was made by Clontech using a combination of random and oligo dt primers. Library was plated and archived by Russell Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

## ORIGIN

Query Match 8.0%; Score 120; DB 2; Length 605;  
Best Local Similarity 52.1%; Pred. No. 7.9e-19;  
Matches 267; Conservative 0; Mismatches 245; Indels 0; Gaps 0;  
QY 415 TGCTGGAGACTATCGCTCTTTATGGCCCTCAAGCCGGTTTGACGCAAGCCCTGG 474  
DB 553 TGGCTTGAAGATGCTGCATTTTTCGCTGCTATCGAATAGCATCAATGAGTGTCTCG 494  
QY 475 AACGAGTGAAGCCCGAGCTGGCGACCGTGAACCGCTGCCCTGGCCAGGCGCGAG 534  
DB 493 TCTGAGTGCCTGAACCGCTGAAGATCGCCATCTCTGGAGCCTTGAAGATATATGAA 434  
QY 535 GAGCTGGCGAGGAGTGGCCCTTTACGAGTGGATTCAAGTGGCTTTTATCTGGAAATGG 594  
DB 433 AACGAGGAGCTTTATAGAAATTTATGGCAACAGTCTTCTATTGAAAGCAATGG 374  
QY 595 GGCCAGACCAAGCCCTATCGCGAATCAAGGGGAGTTCAAGATTATCGGCGATATGCCCATC 654  
DB 373 AAACGGTTCGTTTCAATGCACAGAGCTTGGCATCAGCATATGGTGGTGCATGCCAATA 314

QY	655	TTTGTGGCCCTTGATCTCAGATGCTGGGCCAACCCGCGAGTACTTCTACCTCGAGGCC	714
Db	313	TATGTTGGCTACCAACAGCGCAGATGTTTGGGCAACACAGGAATCATTTTGTCTGGCAAT	254
QY	715	GATGGCAACCCACGGTGGGGGGTTCGGGGGACTTCTCCGAAACCGGCCAG	774
Db	253	AATGGTTTCCCAATGTTGTTAGTGGTTCCTCCTGATGGTTTCAGTAATACTGGTCAG	194
QY	775	CTCTGGGGCAATCCGCTCTATCGCTGGGATGTGATGAAAGGGCAACATTTTGGCTGGTGC	834
Db	193	TTATGGACACGCCATCTGATGACTGGAATCTATGAAGCAGATGCTTTCGATGGTGG	134
QY	835	ATTGCCCGCAATAAGCAGTCTGCTCAAGCAGTGGCCACCTGTGGCCATCGACCACTTCGC	894
Db	133	GTAAGAGGATTAACCGTCCCTTGATTTGATGATGAGTTCCGTTATGACCAATTTCCGC	74
QY	895	GGGTTTGAAGCTACTCGGAGGTTCCGTTGG	926
Db	73	GGGCTTGTGTTTGGGCGAGTTCTTCGGG	42
RESULT 13			
CF013402			
LOCUS			
DEFINITION			
CF013402			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	571	CAGTGGCTTTTATCTGGAATGGGGCCAGACCAAGGCTATGCGGAATCCAAAGGGGATT	630
Db	40	CAGTTCTTATTTCAAGGCAATGGCAGGGATTCGTAATATGACAAAGCTGGTATC	99
QY	631	CAGATTATCGGCGATATGCCCATCTTTGTGGCTTCGATTCCTCAGATGTTCTGGGCCAAC	690
Db	100	ACCATCATGGGTGACATGCTATATATGTTGGCTACCATAGCGCAGATGTTTGGCGCAAC	159
QY	691	CGCAGTACTTCTACCTGAGGCCCGATGGCAACCCACGGTGGTGGCGGGCTTCGGCGG	750
Db	160	AGGAATCATTTTGTCTGCAACAAACGGTTTCCCACTTTTCGTTAGTGGCGTTCCACT	219
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
ORIGIN			
Query Match			

```
Matches 289; Conservative 0; Mismatches 286; Indels 0; Gaps 0;
QY 415 TGGCTGAAGACTATGCGCTCTTTATGCGCCCTCAAGACCCGGTTTGACGGCAAGCCCTGG 474
Db 104 TGGCTTGAAGATGCTGCACTTTTTCAGCTATTGACCATAGCATGTATACATTTCTCTGG 163
QY 475 AACGAGTGGAGCCCGAGCTGCGCGAGCGTGAACCGGCTCCCTGCGCAGGCGCCGTGAG 534
Db 164 ACTGAGTGGCGTGAACCTTTAAAAATCGTCATCTTGATGCTTTGGAAGAAGCATATCGA 223
QY 535 GAGCTGCCGAGGAGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTATCTGGATGG 594
Db 224 ATCCGGAAGACTATATAGACATTTTGTGGCCCAACAATTTTGTTCAAAAACAATGG 283
QY 595 GGCAGACCAAGGCTATGCGGAATCCAAAGGGATTCAGATTTATCGGCGATATGCCCATC 654
Db 284 GTAAACATTCACATCGCGGCACAACTAGGATCAAAATTTATGGTGATATGCCGATT 343
QY 655 TTTGTGGCCTTCGATTCCTCAGATGCTGCGGCCAACCCGAGTACTTCTACCTCGAGGCC 714
Db 344 TATGTTGGTCAATCAAGTGCAGATGTGTGGGCAAAATAAGAAATCCCTTTATGCTGAACAG 403
QY 715 GATGGCAACCCAGCGTGGTGGCGGCTTCGCGGAGCTACTTCTCCGAAACCGGCGCAG 774
Db 404 ATTGGTTATCCAGTTCCTGTCAGTGGCGTTCGCGCTGATTGTTTATGGAACCTGGTCAA 463
QY 775 CTCTGGGCAATCGCTCTATCGTGGGATGTGATGGAGGAGCAACTTTGCTGCTGTC 834
Db 464 TTAGGGGAAGTCTCTATATGATTTGAGGGGCAATGGAAGAAATAGATATGCTTGGTGG 523
QY 835 ATTGCCCGCATAGGAGTGCCTCAAGCAGTGCACCTGGTGGCGCATGCACCACTTCCGC 894
Db 524 ATAAAGAAATAAACGCTCTTGATCTGATGATGAGTTTCGCATAGATCACTTCCT 583
QY 895 GGGTTTGAAGCTACTGCGAGGTTCCGTTTGGCGGCGCCCAATGCTGTGAGGGCGCGTGG 954
Db 584 GGAAGTGTGATGCTGAGTGGCTGCGCTTCGGAAGCAAGTGCCTGCAATGTTTGAAGATGG 643
QY 955 GTCAGAGCCCGAGGAGAGCTGTTGCTGCGGT 989
Db 644 AAGGCTGGACAGGGAACCTTCTTTGATGCTGT 678
```

```
RESULT 15
LOCUS CV302316
DEFINITION CV302316.1 Swollen Stolon Solanum tuberosum cdna clone 67345 5', mRNA
sequence.
ACCESSION CV302316
VERSION CV302316.1 GI:52619649
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
```

```
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Solanum.
1 (bases 1 to 769)
```

```
REFERENCE Flink, B., Rothwell, C., Sardana, R., Griffiths, R., Legue, M.,
AUTHORS DeKoeyer, D., Audy, P., Goyer, C., Li, X.-Q., Wang-Pruski, G. and
Regan, S.
Generation of ESTs from swollen stolon tissues of potato
```

```
TITLE Unpublished (2004)
JOURNAL Contact: Barry Flink
COMMENT The Canadian Potato Genome Project - BioAtlantech
921 College Hill Rd, Fredericton, ON, E3B 6Z9, CANADA
Email: bflink@biatlantech.nb.ca
Seq primer: T3.
```

```
FEATURES Location/Qualifiers
source
1..769
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Shepody"
/db_xref="taxon:4113"
```

```
/clone="67345"
/tissue_type="Stolon"
/lab_host="XL10-Gold"
/clone_lib="Swollen Stolon"
/notes="vector: pBluescript II SK(+) XR; Site: 1: EcoRI;
Site 2: XhoI; supplier: Developmental series. Plants from
pathogen-free Solanum tuberosum var. Shepody, clone 1756,
nuclear stock were grown in a greenhouse under natural
conditions. RNA was isolated from swollen stolon tissue,
3-10mm in length, which was cut from the tip, to the base
of swelling."
```

## ORIGIN

```
Query Match 7.8%; Score 116.6; DB 7; Length 769;
Best Local Similarity 50.1%; Pred. No. 5.8e-18;
Matches 358; Conservative 0; Mismatches 339; Indels 18; Gaps 2;
QY 716 ATGGCAACCCCAAGGCTGCGTTCGCGGCGTCTTCGCGGAGCTACTTCTCCGAAACCGGCGCAG 775
Db 17 AAGGTTTCCCTCTTATAGTTAGTGGTGTTCCTCCAGACGCTTTAGTGAACCTGGTCAAC 76
QY 776 TCTGGGGCAATCGCTCTATCGTGGGATGTGATGGAAGGCAACTTTGCTGCTGCTGCA 835
Db 77 TATGGGGCAGCCCTCTCTATGATTGGAAGCCATGGAAGAGGATGGAATTTTCATGGTGG 136
QY 836 TTGCCCGCATAGGCAAGTGCCTCAAGCAGTGCACCTGGTGGCGCATCGACCACTTCCGCG 895
Db 137 TATGCCGAATTCACGTCGCAACGATCTTTTGTATGTAATTTAGGATAGATCACTTTAGAG 196
QY 896 GGTTTGAAGCTTACTGGGAGTTCGTTTGGCGGCGCCCAATGCTGTGAGGGCGCTGGG 955
Db 197 GATTTGCTGGAATTTGGGCTGTTCCTCTGAGGAAAAAATTTGCAATCTTGGGACGCTGGA 256
QY 956 TCAAGGCCCAAGGAGAGCTGTTGCTGCGGTGGGGCCCCCACTGAGGATGCGGCCA 1015
Db 257 AGGTGGGACCTCGAAAAACCTTTGTTGATGCTATCTTACAAGCTGTGGGAAGATCAATA 316
QY 1016 TCATTGCCGAACACCTGGGGGTGATCACCCCGAGGTGAGGCTTTGGCGCATGGGTTCG 1075
Db 317 TTATAGCAGAGACTTGGGAGTAAATTACCGAGAGCGTTGTTCAAGTAAGAAAGTCCATTG 376
QY 1076 GGTTCGCCGCGCATGAAGATTTTTCAGATTTGCTTTTTCGGTGAGGACAAACGCTTTTTCG 1135
Db 377 AGGCACCTGGAATGGCTGCTACTCCAGTTTTCATTTGGCAGTGCAGCTGAAAAACCTCAT 436
QY 1136 CCCACAACTACCCCGGCGAGCAATGCTGTGATAGCGGAAACCCACGACGACGACA 1195
Db 437 TACCTCACAATCATGAGCA---GAACCAAGTAGTGTATATCTGGAACACATGCAATGATA 493
QY 1196 CCACCTGGGATGGTTTCGCGACCGCGGAGCGCGAGCGGCTTTCATCGCGGCTTACC 1255
Db 494 CGATCGAGGTTGGTGGGATCTTTTCCACAGAGAGAGAAATCCATGCTACTAAAGTATT 553
QY 1256 TGGCCCGCTATGGCATCCGCTGTTTGTGCGAATACAGAGTCCGCGGCGCTTTGATCGAGC 1315
Db 554 TATC-----AAATATTGAGGAGAGGAAATATCATCGGCTTGTATCGAAG 598
QY 1316 TGGCCCTCAAAGCCCGGCGAGCTGGCTATTTGCTCTTTGACGAGACGCTGCTGGGCGTGG 1375
Db 599 GTGCAAGTTCTTCTGTAGCCCGCTATTGCAATATATACCGATGCGAGGATGTTCTTGGGCT 658
QY 1376 GCCCGGAGCGCGCATGAACCTTCCCGGACGCTGGGGAGCAACTGGGCGTGGCG 1430
Db 659 GGAGTGATTCAGAAATGAACATTCAGCAACTGGAAGAACTGGAGTTGAG 713
```

Search completed: January 14, 2006, 14:44:25  
Job time : 5560 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2006, 10:41:10 ; Search time 300 seconds  
(without alignments)  
8905.584 Million cell updates/sec

Title: US-10-618-976-1

Perfect score: 1503

Sequence: 1 atgcaactccaacgcgttt.....ccgaggccagccagcgct 1503

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1503	100.0	1503	3	US-09-687-360-1
2	289	19.2	1518	3	US-09-583-110-1567
3	287.4	19.1	5910	3	US-08-961-527-165
4	265.6	17.7	1377	3	US-09-107-433-336
5	113.2	7.5	1485	3	US-09-902-540-2554
6	113.2	7.5	15377	3	US-09-902-540-1116
7	64.2	4.3	1557	3	US-09-719-978-1
8	47.2	3.1	2930	3	US-09-902-540-9600
9	47.2	3.1	12194	3	US-09-902-540-1091
10	46.4	3.1	1683	2	US-08-838-543-1
11	46.4	3.1	1877	3	US-09-780-173A-10
12	46.2	3.1	11220	3	US-09-105-537-32
13	46.2	3.1	36778	3	US-09-105-537-5
14	46.2	3.1	38506	3	US-09-320-878-19
15	46.2	3.1	38506	3	US-09-141-908-1
16	46.2	3.1	38506	3	US-09-657-440-19
17	46.2	3.1	38506	3	US-09-793-708-19
18	46	3.1	552	3	US-09-003-287-3
19	44	2.9	1776	3	US-09-902-540-7046
20	44	2.9	4532	3	US-09-902-540-626
21	43.2	2.9	2967	3	US-09-185-501B-12
22	42.8	2.8	1287	3	US-09-252-991A-5766
23	42.8	2.8	1329	3	US-09-252-991A-5804
24	42.8	2.8	1857	3	US-09-252-991A-5847

25	42	2.8	666	3	US-09-902-540-9322	Sequence 9322, Ap
26	42	2.8	9839	3	US-09-902-540-996	Sequence 996, App
27	41.8	2.8	5811	3	US-10-152-886-2	Sequence 2, Appli
28	41.4	2.8	699	3	US-09-252-991A-9676	Sequence 9676, Ap
29	41.4	2.8	1062	3	US-09-902-540-3602	Sequence 3602, Ap
30	41.4	2.8	1371	3	US-09-252-991A-9717	Sequence 9717, Ap
31	41.4	2.8	20113	3	US-09-902-540-1173	Sequence 1173, Ap
32	40.8	2.7	1208	2	US-08-403-852D-4	Sequence 4, Appli
33	40.8	2.7	1208	3	US-08-510-646B-4	Sequence 4, Appli
34	40.8	2.7	1208	3	US-09-231-818-4	Sequence 4, Appli
35	40.8	2.7	1208	3	US-09-635-359B-4	Sequence 4, Appli
36	40.8	2.7	5392	2	US-08-403-852D-1	Sequence 1, Appli
37	40.8	2.7	5392	3	US-08-510-646B-1	Sequence 1, Appli
38	40.8	2.7	5392	3	US-09-231-818-1	Sequence 1, Appli
39	40.8	2.7	5392	3	US-09-635-359B-1	Sequence 1, Appli
40	40.6	2.7	360	3	US-09-252-991A-6224	Sequence 6224, Ap
41	40.6	2.7	633	3	US-09-252-991A-6234	Sequence 6234, Ap
42	40.6	2.7	987	3	US-09-758-759-70	Sequence 70, Appl
43	40.6	2.7	1029	3	US-09-252-991A-6399	Sequence 6399, Ap
44	40.6	2.7	1050	3	US-09-252-991A-10218	Sequence 10218, A
45	40.6	2.7	1122	3	US-09-252-991A-9836	Sequence 9836, Ap

## ALIGNMENTS

RESULT 1  
US-09-687-360-1  
; Sequence 1, Application US/09687360  
; Patent No. 6617143  
; GENERAL INFORMATION:  
; APPLICANT: Fukuyama, Shiro  
; TITLE OF INVENTION: Polypeptides Having Glucanotransferase  
; FILE REFERENCE: 6012.200-US  
; CURRENT APPLICATION NUMBER: US/09/687,360  
; CURRENT FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: PA 1999 01501  
; PRIOR FILING DATE: 1999-10-20  
; PRIOR APPLICATION NUMBER: PA 1999 01641  
; PRIOR FILING DATE: 1999-11-15  
; PRIOR APPLICATION NUMBER: 60/160,903  
; PRIOR FILING DATE: 1999-10-22  
; PRIOR APPLICATION NUMBER: 60/166,539  
; PRIOR FILING DATE: 1999-11-18  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1503  
; TYPE: DNA  
; ORGANISM: Thermus rubens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1503)  
US-09-687-360-1

Query Match 100.0%; Score 1503; DB 3; Length 1503;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGCAACTCCAACGGCGCTTTTGGAAATTTTGTCTCCACCCACACAGTTTTCGGGTCGCTGG	60
Db	1	ATGCAACTCCAACGGCGCTTTTGGAAATTTTGTCTCCACCCACACAGTTTTCGGGTCGCTGG	60
Qy	61	GGGATTGGGGCTCTGGCGCGGAGCCGAGCGGTTTTCGACTGGCTGCCGATCGCGGA	120
Db	61	GGGATTGGGGCTCTGGCGCGGAGCCGAGCGGTTTTCGACTGGCTGCCGATCGCGGA	120
Qy	121	GGCCGCTGGTGGCAGGCTTTACCGCTGGGCCCTTACAGTTACGGGACTCGCGGTACCGAG	180
Db	121	GGCCGCTGGTGGCAGGCTTTACCGCTGGGCCCTTACAGTTACGGGACTCGCGGTACCGAG	180
Qy	181	TCCTTCTGGCTTTTGGCCGGTAACCCGTAATTTGGTTGACCCCGAGATGCTGATTGAAAAA	240

181 TCCTTCTCGGCTTTTGGCGGTAAACCGGTATTTGGTTGACCCCGAGATGCTGATGAAAAA 240  
241 GGCTGGCTGGAACAAAGCGAGCGCCCGCGGTATCCGACCCAGCGCGTGGATTATGGC 300  
241 GGCTGGCTGGAACAAAGCGAGCGCCCGCGGTATCCGACCCAGCGCGTGGATTATGGC 300  
301 TGGCTTTACAGACCCCGCTTGGCGGCGCTTTTGGCGGCGTTTTCGGGCAAGG 360  
301 TGGCTTTACAGACCCCGCTTGGCGGCGCTTTTGGCGGCGTTTTCGGGCAAGG 360  
361 GCTTGGCCCAAGNATAGACCCGACTGGAAGCTTTTATCGAGCGCGAGCGCTTCTGGCTG 420  
361 GCTTGGCCCAAGNATAGACCCGACTGGAAGCTTTTATCGAGCGCGAGCGCTTCTGGCTG 420  
421 GAAGACTATGCGTCTTTTATGCGCCTCAAGACCCGCTTTGACGGCAAGCCCTGGAACGAG 480  
421 GAAGACTATGCGTCTTTTATGCGCCTCAAGACCCGCTTTGACGGCAAGCCCTGGAACGAG 480  
481 TGGAGCCCGAGCTGCGGACCGTGAACCCGCTTGCCTTGGCCAGGCGCCGCTGAGGAGCTG 540  
481 TGGAGCCCGAGCTGCGGACCGTGAACCCGCTTGCCTTGGCCAGGCGCCGCTGAGGAGCTG 540  
541 GCGGAGGAGTGGCCCTTTACAGTGGATTTCAGTGGCTTTTATCTGGAATGGGCGAG 600  
541 GCGGAGGAGTGGCCCTTTACAGTGGATTTCAGTGGCTTTTATCTGGAATGGGCGAG 600  
601 ACCAAGCCCTATCGGAATCCAGGGGATTTCAGATTATCGCGATATGCCCATCTTTGTG 660  
601 ACCAAGCCCTATCGGAATCCAGGGGATTTCAGATTATCGCGATATGCCCATCTTTGTG 660  
661 GCTTTCGATCTCAGATGCTGCGGCCAACCCGAGTACTTCTACCTCGAGGCGGATGGC 720  
661 GCTTTCGATCTCAGATGCTGCGGCCAACCCGAGTACTTCTACCTCGAGGCGGATGGC 720  
721 AACCCCAAGTGGTGGCGGGCTTCCGCGGAGTACTTCTCCGAAACGGGCGGCTCTGG 780  
721 AACCCCAAGTGGTGGCGGGCTTCCGCGGAGTACTTCTCCGAAACGGGCGGCTCTGG 780  
781 GGCATCCGCTCTATCGCTGGGATGTGATGAAGGAGCAACTTTGCTGGTGCATTTGCC 840  
781 GGCATCCGCTCTATCGCTGGGATGTGATGAAGGAGCAACTTTGCTGGTGCATTTGCC 840  
841 CGCATAGGCACTGCTCAAGCAGTGGCCACTGCTGGTGGCATCGACCACTTCCGCGGGTTT 900  
841 CGCATAGGCACTGCTCAAGCAGTGGCCACTGCTGGTGGCATCGACCACTTCCGCGGGTTT 900  
901 GAAGCCTACTGGGAGTTCGCTTTGGCGGCCCAATGCTGTGGAGGGGCGCTGGGTCAAA 960  
901 GAAGCCTACTGGGAGTTCGCTTTGGCGGCCCAATGCTGTGGAGGGGCGCTGGGTCAAA 960  
961 GCGCCAGGGAGAGCTGTTGCTGCGGTGCGGCCCAACTGAGCGATGCGCCCATCATTT 1020  
961 GCGCCAGGGAGAGCTGTTGCTGCGGTGCGGCCCAACTGAGCGATGCGCCCATCATTT 1020  
1021 GCGCAAGACCTGGGGGTGATCACCCCGAGGTGGAGGCTTTTGGCGATGGCTTCGGGTTG 1080  
1021 GCGCAAGACCTGGGGGTGATCACCCCGAGGTGGAGGCTTTTGGCGATGGCTTCGGGTTG 1080  
1081 CCGCGCATGAAGATTTTCAGTTTCTGCTTTTCCGGTGAAGCAACGCCCTTTTGGCCAC 1140  
1081 CCGCGCATGAAGATTTTTCAGTTTCTGCTTTTCCGGTGAAGCAACGCCCTTTTGGCCAC 1140  
1141 AACTACCCCGCGCAGGCAATGTGGTGTGTACAGCGGAACCCAGCAACAGCACACACC 1200  
1141 AACTACCCCGCGCAGGCAATGTGGTGTGTACAGCGGAACCCAGCAACAGCACACACC 1200  
1201 CTGGATGGTTCCGCAACCGCGCGGAGCGCGGCTTTCATCGCGGCTTACTTGCGCC 1260  
1201 CTGGATGGTTCCGCAACCGCGCGGAGCGCGGCTTTCATCGCGGCTTACTTGCGCC 1260  
1261 CGCTATGCACTCCGTTGTTGTCGAATACGAGGTTCGGGCGCTTTTGTATCGAGCTGGCC 1320

1261 CGCTATGCACTCCGTTGTTTGTGGAATACGAGTTCGCGGCGCTTTGTATCGAGCTGGCC 1320  
1321 TTCAAAGCCCGCGCAAGCTTGGCTATTTGCTTTTCAGGACGTGCTGGGCTGGGCCCC 1380  
1321 TTCAAAGCCCGCGCAAGCTTGGCTATTTGCTTTTCAGGACGTGCTGGGCTGGGCCCC 1380  
1381 GAGGCCCGCATGAACTTTCCCGGACGCTTGGGGGACAACTGGGCGTGGCGCTACGCCGAA 1440  
1381 GAGGCCCGCATGAACTTTCCCGGACGCTTGGGGGACAACTGGGCGTGGCGCTACGCCGAA 1440  
1441 GCGGACCTCGAGCCCGGCTTGGCCCGGAGACTGCGGCGCCTTGGCGGCGAGCGCGC 1500  
1441 GCGGACCTCGAGCCCGGCTTGGCCCGGAGACTGCGGCGCCTTGGCGGCGAGCGCGC 1500  
1501 GCT 1503  
1501 GCT 1503

## RESULT 2

US-09-583-110-1567  
; Sequence 1567, Application US/09583110  
; Patent No. 6699703

## GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 1567  
; LENGTH: 1518  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-1567

Query Match 19.2%; Score 289; DB 3; Length 1518;  
Best Local Similarity 51.7%; Pred. No. 4.5e-64;  
Matches 764; Conservative 0; Mismatches 695; Indels 18; Gaps 4;

11 AACGCGCTTTTGGAAATTTGCTCCACCCACCAAGTTTTCGGGTGCTGGGGGATTTGGG 70  
8 AAGCTCAAAGTGGTGTGTTGATGCACATCTTCTTCCAGGAGCTTACGGAATCGAT 67  
71 CTCTGGCGCGGAGCGAGCGGTTTGGACTGGCTGGCGGATGGCGGAGCCCGCTGCT 130  
68 CATTTGGTCAAAGTGTCTAGCACTTCGTTGATTTCTTGGTCCGTACAAAACAACGTTACT 127  
131 GGCAGGTCTTACCGCTGGGCGCTTACCACTGACGCGACTCGCGTACAGTCTTCTTCGG 190  
128 GGCATAATCTTCCATATTAGGAGCAACTAGTTACGGGGATTTCTCTTACCAATCTTCTCAG 187  
191 CTTTGGCGGTAAACCCCTATTTGGTTGACCCCGAGATGCTGATTTGAAAAAGCTGGCTGG 250  
188 CTTTCGAGGAGAAACACTCATTTTATCGATTAGATATCTTGGTGGAGCAAGGTTGTTGG 247  
251 AACAAAGCAAGGCGCCCGCGGTATCCGACCCAGCGGTGATTTGGCTGGCTTTACC 310  
248 AAGCAAGTGAACCTTGAAGGAGTTGACTTTGGTAGCGATGCTGCTGAAAGTTGACTATGCTA 307  
311 AGACCGCTGGCCCTGTTGGCGGGCTTTTCGGGGGCTTTTCGGGCAAGGCTTCGGGCC 370  
308 AAATCTACTATGACGCTGCTCTCTTTTAGAAAAAGCGGTGAACGTTCTTTTGAAGTCG 367  
371 AGGATAGACCGGACTCGAAGCCCTTTATCGAGGCGCGAGCGCTTCTGGCTGGAAGCATATG 430



Db 368 GAGATGTTAAAGATTTTGGAGAAATTTGCTCAAGACAACCAATCATGCTTGAGCTCTTTG 427  
Qy 431 CGCTCTTTATGGCCCTCAAGACCGGTTTGAAGCGCAAGCCCTGGAACGAGTGG---AGCC 487  
Db 428 CTGAGTATATGGCTATCAAGAGCATTTTGAACAATCTTGTCTTGACTGAATGGCCAGATG 487  
Qy 488 CCAGAGCTCGCGACCGGTGAACCGGCTGCCCTGGCCAGGCGCCGTGAGGAGCTGCCGAGG 547  
Db 488 CAGATGCTGCTGCTGTAAGCTTCAGACATTTGAAGCTATCGTAGCAATTTGCGACACA 547  
Qy 548 AGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTATCTGGAATGGGGCCAGACCAAGG 607  
Db 548 AGTTGGTTTACCAACCGGTGACTCAATACTCTCTTCCAAACAATGGTTGAATTTGAAG 607  
Qy 608 CCTATGCGATCAAGGGGATTCAGATTATCGGGGATGCCCATCTTTGTGGCCTTCG 667  
Db 608 CTTACGCTTAACGACAACACATCGAATCGTTGGGGACATGCCAATCTACGTAGCGGAAG 667  
Qy 668 ATTCTCTCAGATGCTGGGCCCAACCGCAGTACTTCTACCTCGAGCGGATGGCAACCCCA 727  
Db 668 ATTCAAGTGATATGGGCAATCCACATCTCTTCAAAACAGATGTCAATGGTAAAGCTA 727  
Qy 728 CGGTGGTGGCGGCGTTTCCGGGAGTACTTCTCCGAACCGGCCAGCTCTGGGGCAATC 787  
Db 728 CTTGATCGCAGGATGCCACCATGAGTTTTCTGTAACCTGGTCAGCTTTGGGGTAATC 787  
Qy 788 CGCTCTATCGTGGGATGTGATGTAAGGAGCAACTTTGCTGCTGATGCTGCCCGGATAA 847  
Db 788 CAATCTATGACTGGGAAGCAATGGAACAAGCGGTACAAATGGTGGATTTGAACGCTTGC 847  
Qy 848 GGCAGTGGCTCAAGCAGTGCACCTGCTGGCATCGACCACTTCCGGCGGTTTGAAGCCT 907  
Db 848 GTGAAGCTTCAAAATCTACGATATCGTTCTGATCGACCACTTCCGTGGCTTGAATCTT 907  
Qy 908 ACTGGAGGTTTCCGTTTGGCGGCGCCCAATGCTGTGGAGGGCGCTGGGTCAAAAGCCCG 967  
Db 908 ACTGGGAATCCCTGCTGTTCCGATACAGCAGCACCTGGTGGTGAAGAGTCCAG 967  
Qy 968 GGGAGAGCTGTTGCTGGGTCGGGCGGCCCACTGACGATCGGCCCATATGTCGCAAG 1027  
Db 968 GCTACAAGCTTTTTCAGCGGTTTAAAGGAAGAACTTGGTGAGCTAAACATCATCGCAAG 1027  
Qy 1028 ACCTGGGGTGATCACCCCGAGTGAGGCTTTGGCGGATGCTTCCGGGTCCCGGCA 1087  
Db 1028 ACCTGGCTTCATGACAGATGAAGTATCGAATTCGCTGAACGACTGGCTTCCAGGA 1087  
Qy 1088 TGAAGATTTTGCAGTTTGTCTTTTCCGCTGAGGACACGCTTTTCCGCCCAACACTACC 1147  
Db 1088 TGAAGATTTTCAATTTGCTTCAACCCAGAGAGCAAGCATTTGATAGCCACACT--- 1144  
Qy 1148 CCAGCGCAGCAATGTTGGTGTGACAGCGGAACCCACGACGACACCAACCCCTGGAT 1207  
Db 1145 TGGCACCCTGCTAACTCAGTTATGTACACAGAAACACACGATATAAATACGCTTCTGGTT 1204  
Qy 1208 GGTTCGCAACCGCGGAGCGGCGGCGCTTATGCGGCGCTTACCTGGCGCGCTATG 1267  
Db 1205 GGTACCGTAATGAGATGATGATGCG-----ACTGCTGAGTACATGGCTCGTTACA 1255  
Qy 1268 GCATCGCTTGTTCGGAATAACAGGTTCGGGCGCTTTGATCGAGCTGGCTTCAAAA 1327  
Db 1256 CGAACCTGAAGATACGAACAGTGT---ACAGGCTATGCTTCGTACAGTATTTTCA 1312  
Qy 1328 GCCCGGCAAGCTGGCTATTTGCTTTGAGGAGCTGCTGGGGCTGGGCGCCCGAGGCC 1387  
Db 1313 CAGTTAGCTTTATGGCAATTTGCAATATGCAAGATTTTACTAGAAATTTGATGAGCGACTC 1372  
Qy 1388 GCATGACTTCCCGGAGCGCTGGGGAACAATCGGGGCTGGCGCTACGCCGAGAGCGACC 1447  
Db 1373 GTATGAATTTTCCCATCTTACCTTTGGTGAACATGGTCTTGGCGTATGACTGAAGATCAAT 1432  
Qy 1448 TCGAGCCCGCTGCTGGCGCGGAGCTGGCGGCGCTTGGC 1484  
Db 1433 TGACACCAAGCTGTCGAGGAAGTTTGTGCTGACTTGAC 1469

## RESULT 3

US-08-961-527-165  
; Sequence 165, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 165:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5910 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-527-165

Query Match 19.1%; Score 287.4; DB 3; Length 5910;

Best Local Similarity 51.7%; Pred. No. 1.9e-63;

Matches 763; Conservative 0; Mismatches 696; Indels 18; Gaps 4;

Qy	11	AACGCGCTTTTGGAAATTTGCTCCACCCACACAGTTTTCGGGTGCTGGGGGATTGGGG	70
Db	108	AACGTCAAAGTGTGTGTGATGCATCTCTTCTTCCAGGAGCTTACGGAATCGAT	167
Qy	71	CTCTGGCGCGCAGGCGGAGCGGTTTTTGACTGGCTGGCCGATGGGGAGCCCGCTGGT	130
Db	168	CATTTGGTCAAAGTGTCTAGACTTCGTTGATTTCTTGGTCCGTACAAAACAAGTTACT	227
Qy	131	GGCAGGTCTTACGCTGGGCGCTACAGTTACGGCGACTCCCGTACCAGTCTTCTCGG	190
Db	228	GGCAATCTCTCCATTTAGGAGCAACTAGTTTACGGGATTTCTCTTACCAATCTTCTCAG	287
Qy	191	CTTTTTCGGCTAACCCGTTATTTGGTTGACCCCGAGATGCTGATGAAAAGGCTGGCTGG	250
Db	288	CCTTCGACGAACAACATCTTTATTCGATTTAGATCTTGGTGGAGCAAGTTTGTGG	347
Qy	251	AACAAAGCAAGCGCCCGCCGCTATCCGACCCAGCGCTGGATTATGGCTGGCTTTACC	310
Db	348	AAGCAAGTACCTTGAAGAGTTGACTTTGGTAGCGATGCTCTGAAGTTGACTATGCTA	407
Qy	311	AGACCGCTGGCGCCCTTGTGGCGGGCTTTTCGGGGGTTTCGGGCAAGGGCTTCGGCC	370
Db	408	AAATCTACTATGCACGTCGCTCTCTTTTAGAAAAAGCGGTGAACCGTTCTTTTGAAGTCG	467
Qy	371	AGGATGAAGACCGCACTGGGAAGCCTTTATCGAGGCCGCGGCTTCTGGCTGGAAGACTATG	430

Db 1533 TGACACCGACTGTCGAGGAAGGTTTGCTTGACTTGAC 1569

RESULT 4  
US-09-107-433-336  
; Sequence 336, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,433  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/ 085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-011  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 336:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1377 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...1377  
; SEQUENCE DESCRIPTION: SEQ ID NO: 336:  
US-09-107-433-336

Query Match 17.7%; Score 265.6; DB 3; Length 1377;  
Best Local Similarity 52.1%; Pred. No. 4.2e-58;  
Matches 698; Conservative 0; Mismatches 624; Indels 18; Gaps 4;

Qy 148 GGCCCTACGATTACGGGACTCGCGTACCACTCTTCGCGCTTTTCGCGTACCCG 207  
Db 4 GGAGCAACTAGTTACGGGATTCCTCTTACCAATCTTCTCAGCCTTCGAGGAACACT 63  
Qy 208 TATTGGTTGACCCCGAGATGCTGATTGAAAGGCTGGCTGGAACAAAGCGGCC 267  
Db 64 CATTTATCGATTAGATATCTTGGTGGAGCAAGTTGTTGGAACAGTGCCTTGAA 123  
Qy 268 CCGCGGTATCCGACCCGCGGTGGAATTATGGCTGGCTTTACAGACCGCTGGGCCCTG 327  
Db 124 GGAGTTGACCTTTGAGTAGCGATGCTGTAAGTTGACTATGCTAAAATCTACTATGCACT 183



328 TTGCGGGGGCTTTTGGGGGTTTGGGCAAGGGCTTCGGCCAGGATAAGACCCGACTG 387  
Db CGTCTCTTTTAGAAAAGCGGTGAAAGCTTTCTTGAAGTCGAGATGTTAAAGATTTT 243  
388 GAAGCCTTTATCGAGGCGAGCGCTTCGTGGCTGGAAGACTATGCGCTCTTTATGGCCCTC 447  
Db GAGAAAATTTGCTCAAGAACAACTCATGCTTGTGAGCTCTTGTGAGTATATGGCTATC 303  
448 AAGACCCGGTTTGAAGCGCAAGCCCTGGNACAGATGGAGC---CCGAGCTCGCGAGCGT 504  
Db AAAGAGCAATTTTGAACAATCTTGTGTGAGCTGAATGGCCAGATGAGATGCTGCTGCTCGT 363  
505 GAACCGGCTGCTCGGCGAGCGCCGCTGAGAGCTGGCCGAGGAGGTGGCCCTTTACGAG 564  
Db AAGCTTCAGACTTGAAGCTATCGTGAGCAATGGCAGACAAGTTGGTTTACCAAGCT 423  
565 TGGATTGAGTGGCTTTTATCTGGAATGGGGCGAGACCAAGGCTATGCGGAATCAAG 624  
Db GTGACTCAATACTTCTTCTTCCAAATGGTTGAAATTTGAAGCTTACGCTTAACGCAAC 483  
625 GGGATTGAGTATCGGGGATATGCCATCTTTGTGGCTTCGATTCCTCAGATGTCGG 684  
Db CACATCAAAATCGTTGGGACATGCCAATCTACGTAGCGGAAGATTCAAGTGATATGG 543  
685 GCCAACCCGAGTACTTCTACCTCGAGCGCGATGCGCAACCCAGCTGGTGGGGGGT 744  
Db GCAATCCAACTCTTTCAAAACAGATGTCAATGGTAAGGTACTTGTATCGAGGATGC 603  
745 CCAGGGGACTTCTTCGAAACCGGCGAGCTTCGGGGCAATCGCTCTATCGCTGGGAT 804  
Db CCACCAAGATGATTTCTGTAACTGCTCAGCTTTGGGGTAATCCAACTATGACTGGAA 663  
805 GTGATGGAAGGACAACTTTGCTGCTGATGCTGCGGCAATAGGCAAGCTCGCTCAGCAG 864  
Db GCAATGGAACAAGACGCTCAAAATGGTGAATGAACGCTTGGCTGAAAGCTTCNAATC 723  
865 TGCACCTGCTGCGATCGACCACTTCGCGGGTTTGAAGCTACTGGGAGGTTCGGTT 924  
Db TAGATATGCTTGTATCGAACCACTTCGCTGGCTTGAATCTTACTGGGAAATCCCTGCT 783  
925 GGCAGGCGCAATGCTGTGAGGGGCGCTGGGTCAAAGCCCGAGGGGAGAGCTGTTGCT 984  
Db GGTTCGATACAGCAGCACCTGCTGTGAGTGGGTGAAGGTCAGGCTACAAAGCTTTTGCA 843  
985 GCGGTGGGGCCCACTGAGCGATGCGCCATCATTTGCCAAGACCTGGGGGTGATCACC 1044  
Db GCCGTTAAGGAAGAACTTGTGAGCTAAACATCATCGAGAAGACCTTGGCTTCATGACA 903  
1045 CCGAGGTGAGGCTTTGCGCGATGCTTCGGGTTCCCGGATGAAGATTTTGCAGTTT 1104  
Db GATGAAGTGAATGCAATTTGCGTGAACGTAATGCTTCCAGGAATGAAGATTTCTCAATTT 963  
1105 GCTTTTTCGGGTGAGGACAAAGCGCTTTTGGCCCACTACCCCGGCGAGCAATGTG 1164  
Db GCCTTCAACCCAGAGAGCAAGCAATGATAGCCCACT---TGGCACCTGCTAATCA 1020  
1165 GTGGTGAAGGAAACCGAGCAACAGCAACCACTGGGATGTTCCGCAACCGGCGG 1224  
Db GTTATGTACACAGAAACACAGATAAATAACGGTCTTGTGTTGTTACCGTAATGAGATT 1080  
1225 GAGGCGAGCGGCTTATGCGGGCTTACGCGGCTTACGCGGCTATGCGATCGCTTTGTCG 1284  
Db GATGATGCG-----ACTGTGATGATGATGCTGTTTACAGAACCGCTAAAGAAATAC 1131  
1285 GAATAGAGTTCGGGCGCTTTGATCGAGCTGCGCTTCAAAAGCCCGGCAAGCTGGCT 1344  
Db GAAACAGTGT---ACAGCTATGCTTCGTACAGTATTTTCACTAGTTAGCTTTATGGCA 1188  
1345 ATTGTGCTTTGAGGAGCTGCTGGGCTGGGCGCGGCGGCGGATGAATCTCCCGGA 1404  
Db ATTGCAACTATGCAAGATTTACTAGAATTGGATGAGGAGCTCGTATGAATCTTCCATCT 1248  
1405 CGGCTGGGGGACAACTGGGGCTGCGGCTACGCGGAGGCGACCTTCGAGCCCGCTCTGGCC 1464

Db 1249 ACCCTGGTGAAGAACTGGTCTTGGGTATGACTGAAGATCAATTGACACCAAGCTGTGAG 1308  
Qy 1465 GCGGACTGCGGGCCCTGGG 1484  
Db 1309 GAAGGTTTGTGCTTGAAGTTGAC 1328

RESULT 5  
US-09-902-540-2554  
; Sequence 2554, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barty S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 2554  
; LENGTH: 1485  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
; US-09-902-540-2554

Query Match 7.5%; Score 113.2; DB 3; Length 1485;  
Best Local Similarity 53.8%; Pred. No. 4.2e-19;  
Matches 269; Conservative 0; Mismatches 213; Indels 18; Gaps 1;

Qy 399 CGAGGCGGAGCGCTTCGTGCTGGAAGACTATGCGCTCTTTATGGCCCTCAAGACCCGCTT 458  
Db 378 CGAGCGCCAGGCGCAATGGCTGAGAGCTACGCGCTCTTCCGCCATCAGCGAGCAGGA 437  
Qy 459 TGACGCAAGCCCTGGAACGAGTGGAGCCCCGAGCTGCGGACCGTGAACCGGCTGCCCT 518  
Db 438 AGAGCGCCCGCGCTGCTGGAGTGGCCGAGCGCTGCGCACCGCTCAACCCGAGGCGCT 497  
Qy 519 GGCAGCGGCGCGTGAAGAGCTGGCGCGAGAGTGGCGCTTTACGAGTGAATTCAGTGGCT 578  
Db 498 GGCAGAGAGTCCCGGAGCTGAGCGCGCGCTGCGTACCAAGCTTCCAGTGGT 557  
Qy 579 TTTTATCTGGAATGGGCGAGACCAAGGCTATGCGGAATCAAGGGGATTCAGATTAT 638  
Db 558 GGCAGGAGCAGCTGGAACGAGTGGAGCGGAGCGCGCCGAGGACGCTCTCTCTGTG 617  
Qy 639 CGCGATATGCCATCTTTGTGGCTTCGATTCCTCAGATGCTGGGCCAACCCGCACTA 698  
Db 618 CGCGCAGAGCGCTTTCATCATCGTCAAGCAGCGCGAGCTGTGGCGCACCCCGGACAT 677  
Qy 699 CTTTCTACCTCGAGGCGGATGGCAACCCAGCGTGTGGCGGGCGTTCGCGGCACTACTT 758  
Db 678 CTTGCGCGGAGCGCT-----CGCTGGGCTGCTCGGATGACTT 719  
Qy 759 CTCGGAACCGGCGAGCTCTTGGGGCAATCGCTCTATCGCTGGGATGTGATGAAGGGA 818  
Db 720 CTCGCGCAACCGGCGAGCTTGGGGCTGCGCTTCTACTTTCGACCTGGCGCCATGAGAAGGA 779  
Qy 819 CAATTTGCTGCTGATTTGCCCGCATAGGAGTGCCTCAAGCATGCGCCCTGGTGG 878  
Db 780 CGACTTTCGCTGCTGCTGAAGACGCGCGGAGGCGGCGGCTACTTACGACTTGGCGCG 839  
Qy 879 CATCGACCACTTCCGCGGCT 898  
Db 840 CGTGAACCAAGCGGCTGGGT 859

RESULT 6  
US-09-902-540-1116/c

```

; Sequence 1116, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1116
; LENGTH: 15377
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1116

Query Match 7.5%; Score 113.2; DB 3; Length 15377;
Best Local Similarity 53.8%; Pred. No. 9.6e-19;
Matches 269; Conservative 0; Mismatches 213; Indels 18; Gaps 1;

Qy 399 CGAGCGCAGCGCTTCTGGCTGGAGACTATGCGCTCTTTATGCGCTCAAGACCCGGTT 458
Db 8803 CGACGCCAGCGCGAATGGCTGGAGAGCTACGCGCTCTTCAACCGCCATCAGCGAGCAGGA 8744

Qy 459 TGACGGCAAGCCCTGGAAACAGTGGAGCCCGAGCTGCGGACCGTGAACCGGCTGCCCT 518
Db 8743 AGACGGCGCCCGTGGTGGAGTGGCGGAGCCCTGCGGACCCGCTCAACCGGAGCCCT 8684

Qy 519 GGCCAGGCGCCGCTGAGGAGCTGGCGGAGAGGTGGCCCTTTACGAGTGGATTCAGTGGCT 578
Db 8683 GGCCAGAGTCCCGGAGAGCTGGAGCGCGCGTGGCTTACACGCTGGCTCCAGTGGGT 8624

Qy 579 TTTTATCTGAATGGGGCGAGCAGAGCGCTATGCGGATCCGATCCAAAGGGATTCAGATTAT 638
Db 8623 GGCCAGCAGCAGTGGAAACAGAGTGGCGACGCGGCGCGCGCCCAAGGACGTCCTCCTGTG 8564

Qy 639 CGCGGATATGCCATCTTGTGGCTTTCGATTCCTCAGATCTCTGGCCCAACCGCAGTA 698
Db 8563 CGCGCAGAGCGCTTTCATCTCGTACGAGCAGCGCGGAGCTGTGGCGCACCCGGACAT 8504

Qy 699 CTCTACCTCGAGCGCGATGAGCAACCCCAACGCTGTGGCGGCGCTTCGCGGAGTACTT 758
Db 8503 CTTGCGGCGCAGCGC-----CGCCTGGCGCTGGCTCCGATGACTT 8462

Qy 759 CTCCGAAACCGCGCAGCTCTGGGGCAATCCGCTCTATCGCTGGGATGTGATGGAAGGA 818
Db 8461 CTCCGCCACGCGCGCAGGACTGGGCGCTGCGCTTCTCGACTTGGCGGCGCATGGAAGGA 8402

Qy 819 CAATTTGCTGTGTCATTTGCCCATAGGACAGTTCGCTCAAGCAGTGCACCTGGTGGC 878
Db 8401 CGACTTCGCTGTGTGAAGACGCGCGGAAGAGCGGCGCCAGCTACTACGACTTGGCGCG 8342

Qy 879 CATGACCACTTCGCGCGGT 898
Db 8341 CTGACCAACCGGCTGGGGT 8322

RESULT 7
US-09-719-978-1
; Sequence 1, Application US/09719978
; Patent No. 6858717
; GENERAL INFORMATION:
; APPLICANT: Biogen
; TITLE OF INVENTION: Method for obtaining modified polysaccharides
; FILE REFERENCE: BFF 98/0267
; CURRENT APPLICATION NUMBER: US/09/719,978
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: FR/9807589
; PRIOR FILING DATE: 1998-06-16

; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1557
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii
US-09-719-978-1

Query Match 4.3%; Score 64.2; DB 3; Length 1557;
Best Local Similarity 53.8%; Pred. No. 1.5e-06;
Matches 157; Conservative 0; Mismatches 133; Indels 3; Gaps 1;

Qy 49 CCGGTCGCTGGGGATTTGGGCTCTGGGCGCGAGCGCGAGCGGTTTTTGGACTGCTG 108
Db 1 CCGGGTGGCTACGGTATTTGGGATTTGGTATGAGGCCAAGCGCTTCGTCGACTGCTC 60

Qy 109 GCCGATGCGGGAGCCCGCTGGTGGCAGGTCTTACCGTGGGCGCTTACAGTTACGG--C 165
Db 61 GCCGACCAACGGCATGCGAGTCTGGCAGCTGCTCGCTGGTGGTGGCGGACCCCATGTAC 120

Qy 166 GACTCGCGTACCAGTCTCTCTCGGCTTTTCCGCTAACCCGATTTTGGTTGACCCGAG 225
Db 121 TACTCCCCCTACTCGGGCAGCGGCAACTCGCGGCAACCCCTTGTGTGAGCATTTGAG 180

Qy 226 ATCTGATTTGAAAAGCGCTGGCTGGAAACAAAGCGAGCGCGCGCGCTATCCGACCCAG 285
Db 181 GAGCTCATCAGACGAGGACTGCTGGAGTTCTCGGAGAGCGCGCGCGCTGCCATCGCC 240

Qy 286 CGGCTGGATTTATGGCTGGCTTTTACAGAACCGCTGGCGCTTGTGGCGCGCGC 338
Db 241 GAGTTTGACTATCCCGCTGTGGCGCGGCGCAAGCTGCCCTGTCTCAAGCGCGC 293

RESULT 8
US-09-902-540-9600
; Sequence 9600, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9600
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-9600

Query Match 3.1%; Score 47.2; DB 3; Length 2930;
Best Local Similarity 51.5%; Pred. No. 0.041;
Matches 134; Conservative 0; Mismatches 123; Indels 3; Gaps 1;

Qy 1218 CGCGCGGAGCGCGAGCGGCGCTTTCATGCGGCGCTTACCTGGCGCGCTATGGCATCCGTTG 1277
Db 720 CGCGCGCTGGCGAGCGCTGTCTGGGCAAGTCTGTCAACGCGGCGCAGAGCTGCGT 779

Qy 1278 TTTGTCGGAATACAGAGTCCGGCGCTTTTGATCGAGTGGCTTCAAAAGCCGCCCAA 1337
Db 780 GCGCGCGGACTACGTGTTCTGTCACGCGTCAAGAGGAGCGGCGCTTCTTGGACGCGCTCAA 839

Qy 1338 GCTGGCTATTGCTCTTTGAGGAGCGTCTGGGCTGGCGCGCGCGCGCGCATGAACTT 1397
Db 840 GCGGCTCATCAGCGCTTCTACGCGCGCACCGAGCAGGAGCGGCGGAGCGGCGCTGACCT 899

Qy 1398 CCGCGGAGCGCTGGGCGCACTGCGGCTGGCGTACGCCGAGGCGCACTTCGAGCGCGG 1457
Db 1398 CCGCGGAGCGCTGGGCGCACTGCGGCTGGCGTACGCCGAGGCGCACTTCGAGCGCGG 1457
```

Db 900 GACGCGA---CTGGTGGACCGCGTGGCGGTGAAGGACCTGCTCAAGCGCTC 956  
Qy 1458 TCTGCGCGGCGGACTGCGGG 1477  
Db 957 GGTGGCGCGGAGCGGAAGG 976

## RESULT 9

US-09-902-540-1091  
; Sequence 1091, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 1091  
; LENGTH: 12194  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)...(12194)  
; OTHER INFORMATION: unsure at all n locations  
US-09-902-540-1091

Query Match 3.1%; Score 47.2; DB 3; Length 12194;  
Best Local Similarity 51.5%; Pred. No. 0.068;  
Matches 134; Conservative 0; Mismatches 123; Indels 3; Gaps 1;

Qy 1218 CGCGCGGAGCGCGGCGCTTCATCGGGCTACTGGCGCTATGGCGTATGGATCGGTTG 1277  
Db 2146 CGCGCGGCTGCGGAGCGCTGTGCTGGGCAAGTTCTCAACGGCGCGGCGGCGT 2205  
Qy 1278 TTGTGGAATACAGAGTCGCGGCGCTTTGATCGAGCTGGCTTCAAAAGCCGCGCAA 1337  
Db 2206 GCGCGGAGTACTGTGTCGACGCGTGAAGAGCGGCGTCTGACGCGCTCAA 2265  
Qy 1338 GCTGGCTATTGTGCTTTGAGGACGTCGTGGGCTGGGCGCCGAGCGCCGATGAATT 1397  
Db 2266 GCGCTCCATCACGCGCTTCTACGCGCGCACCGAGCAGCGCGGAGCGCTGACCT 2325  
Qy 1398 CCAGGAGCGCTGGGAGCAACTGGGCGTGGCGTACCGGAGGCGGAGCTGAGCGCGG 1457  
Db 2326 GACGCGA---CTGGTGACCGGTGGCGTGGCGGTTGAAGGACCTGCTCAAGCGCTC 2382  
Qy 1458 TCTGGCGCGGAGTGGGG 1477  
Db 2383 GGTGGCGCGGAGCGGAAG 2402

## RESULT 10

US-08-838-543-1  
; Sequence 1, Application US/08838543  
; Patent No. 5994623  
; GENERAL INFORMATION:  
; APPLICANT: KREBBERS, ENNO  
; APPLICANT: BROGLIE, KAREN E.  
; TITLE OF INVENTION: CORN 4-(GLUCANOTRANSFERASE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: UNITED STATES OF AMERICA

ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.50 INCH DISKETTE  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WINDOWS 95  
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (VER. 7.0)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/838,543  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: MAJARIAN, WILLIAM R.  
REGISTRATION NUMBER: P-41,173  
REFERENCE/DOCKET NUMBER: BB-1101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-992-4926  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1683 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..1489  
US-08-838-543-1

Query Match 3.1%; Score 46.4; DB 2; Length 1683;  
Best Local Similarity 53.3%; Pred. No. 0.054;  
Matches 98; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy 737 CGGCGCTTCGCGGAGTACTTCTCCGAACCGCGCAGCTCTGGGGCAATCGGCTCTATC 796  
Db 171 CTGGAGCGCTCTCGATTATTTTGACAAGATGACAAAATTTGGGGTTTCTTACATATA 230  
Qy 797 GCTGGGATGTGATGGAAGGACAACTTTCCTGTCATTCGCCGATAGGCGTGC 856  
Db 231 ACTGGGAGGAGATGTCAAAGGATAATTTATGGGTGGTGGGAGCTCGTCTGACACATGG 290  
Qy 857 TCAAGCAGTGCACCTGCTGGTCATCGACCTTCCGCGGTTTGAAGCCTTACTGGGAGG 916  
Db 291 CAAAGTCTTCACAGCATACAGATAGACCAATCTTGGTTCTTTAGGATATGGAGC 350  
Qy 917 TTCC 920  
Db 351 TTCC 354

## RESULT 11

US-09-780-173A-10/c  
; Sequence 10, Application US/09780173A  
; Patent No. 6455307  
; GENERAL INFORMATION:  
; APPLICANT: Robert McKay  
; APPLICANT: Susan M. Preier  
; APPLICANT: Jacqueline Wyatt  
; TITLE OF INVENTION: ANTISENSE MODULATION OF CASEIN KINASE 2-ALPHA PRIME EXPRESSION  
; FILE REFERENCE: RTS-0165  
; CURRENT APPLICATION NUMBER: US/09/780,173A  
; CURRENT FILING DATE: 2001-02-08  
; NUMBER OF SEQ ID NOS: 95  
; SEQ ID NO 10  
; LENGTH: 1877  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (344)...(1396)  
US-09-780-173A-10

Query Match 3.1%; Score 46.4; DB 3; Length 1877;

Best Local Similarity 53.3%; Pred. No. 0.057; Mismatches 98; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 1312 GAGTGGCTTCAAAAGCCCGGCAAGCTGCTATTGCTTTTCAGAGACGTGCTGGG 1371  
 Db 436 GAGTGGCTTGTAGTCCAGTACTCGCGCTCTCAGGCTGTTCACCTCGCGTAGAC 377

QY 1372 CTGGGCCCCGAGGCGCCGATGAATCTCCCGGACGGCTGGGGGACAACTGGGGCTGGCGC 1431  
 Db 376 CCGGGCCGACTGCG 317

QY 1432 TAGCCGAAGCGACCTCGAGCCCGGTCTGCGCGCGGGACTGCGGGCCCTGGCGAGGCC 1491  
 Db 316 GCGCGCGAGGCG 257

QY 1492 AGCC 1495  
 Db 256 GGCC 253

RESULT 12  
 US-09-105-537-32  
 ; Sequence 32, Application US/09105537A  
 ; Patent No. 6265202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D. H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600.438US1  
 ; CURRENT APPLICATION NUMBER: US/09/105.537A  
 ; CURRENT FILING DATE: 1998-06-26  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 32  
 ; LENGTH: 11220  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 US-09-105-537-32

Query Match 3.1%; Score 46.2; DB 3; Length 11220;  
 Best Local Similarity 47.4%; Pred. No. 0.12;  
 Matches 138; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 1211 TCCGCACCGCGCGAGCGCGGCTTTCATGCGGGCTTACTGCGGCTTATGCGGCTATGGCA 1270  
 Db 9737 TCGCGCGCGCGCGAGCCCTTCCGACGTACGCGAGCCCGCCACACGCGGCAAGTGG 9796

QY 1271 TCCGCTTGTTCGGAATACGAGTTCGCGGCGCTTTGATCGAGTGGCTTCAAAGCC 1330  
 Db 9797 TCCTCAGATGCGTCCGCGCTCGACCGGAGGTACGCTCTGCTGACCGCGCGCACCG 9856

QY 1331 CGGCAAGCTGCTATTGCTTTGAGGAGTCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 1390  
 Db 9857 GTGCGCTGGGGGACATCGTGGCGCGGACGTGTTGGGCGAGTGGGGCGGTACGAGCCTGC 9916

QY 1391 TGAACCTCCCGAGCGCTGGGGGACAACTGGGCTGGGCTGACCCGAAAGCGACCTCG 1450  
 Db 9917 TGCTGTGAGCGCGCGGGGACGAGACGCCCGCGGGCGCGGCGGTGTCGACGAGTGG 9976

QY 1451 AGCCCGGTCTGCGCGCGGACTGCGGGGCTTGGCGGCTTGGCGGCGGCGGCGGCGG 1501  
 Db 9977 AGGCTCTGGAGCCGACGCTCTCGTGGCGCGGTGCGGAGTGGCGGCGGCGGCGG 10027

RESULT 13  
 US-09-105-537-5  
 ; Sequence 5, Application US/09105537A  
 ; Patent No. 6265202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D. H.  
 ; APPLICANT: Liu, H.

APPLICANT: Xue, Y.  
 APPLICANT: Zhao, L.  
 TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 FILE REFERENCE: 600.438US1  
 CURRENT APPLICATION NUMBER: US/09/105.537A  
 CURRENT FILING DATE: 1998-06-26  
 NUMBER OF SEQ ID NOS: 43  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 5  
 LENGTH: 36778  
 TYPE: DNA  
 ORGANISM: Streptomyces venezuelae  
 US-09-105-537-5

Query Match 3.1%; Score 46.2; DB 3; Length 36778;  
 Best Local Similarity 47.4%; Pred. No. 0.18;  
 Matches 138; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 1211 TCCGCACCGCGCGAGCGGCTTTCATGCGGCTTACTGCGGCTTATGCGGCTATGGCA 1270  
 Db 25424 TCGCGCGCGCGCGAGCGCTTCCGCGACGTACGCGAGCCCGCCACACGCGGCAAGTGG 25483

QY 1271 TCCGCTTGTTCGGAATACGAGTTCGCGGCGCTTTGATCGAGTGGCTTCAAAGCC 1330  
 Db 25484 TCCTCAGATGCGTCCGCGCTCGACCGCGAGGTACGCTCTGCTGACCGCGGCGCACCG 25543

QY 1331 CGGCAAGCTGCTATTGCTTTGAGGAGTCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 1390  
 Db 25544 GTGCGCTGGGGGACATCGTGGCGCGGACGTGTTGGGCGAGTGGGGCGGTACGAGCCTGC 25603

QY 1391 TGAACCTCCCGAGCGCTGGGGGACAACTGGGCTGGGCTGACCCGAAAGCGACCTCG 1450  
 Db 25604 TGCTGTGAGCGCGCGGGGACGAGACGCCCGCGGCGGCGGAGTCTGTCGACGAGTGG 25663

QY 1451 AGCCCGGTCTGCGCGGAGTTCGCGGCTTGGCGGCTTGGCGGCTTGGCGGCGGCGGCGG 1501  
 Db 25664 AGGCTCTGGAGCGCGACGTCTCGTGGCGCGGTGCGGAGTGGCGGCGGCGGCGG 25714

RESULT 14  
 US-09-320-878-19  
 ; Sequence 19, Application US/09320878A  
 ; Patent No. 6117659  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ASHLEY, Gary  
 ; APPLICANT: BETLACH, Melanie C.  
 ; APPLICANT: BETLACH, Mary C.  
 ; APPLICANT: MCDANIEL, Robert  
 ; APPLICANT: TANG, Li  
 ; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
 ; FILE REFERENCE: 30062202120  
 ; CURRENT APPLICATION NUMBER: US/09/320.878A  
 ; CURRENT FILING DATE: 1999-05-27  
 ; EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
 ; EARLIER FILING DATE: 1998-08-28  
 ; EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
 ; EARLIER FILING DATE: 1998-05-06  
 ; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
 ; EARLIER FILING DATE: 1997-04-30  
 ; EARLIER APPLICATION NUMBER: 60/119,139  
 ; EARLIER FILING DATE: 1999-02-08  
 ; EARLIER APPLICATION NUMBER: 60/100,880  
 ; EARLIER FILING DATE: 1998-09-22  
 ; EARLIER APPLICATION NUMBER: 60/087,080  
 ; EARLIER FILING DATE: 1998-05-28  
 ; NUMBER OF SEQ ID NOS: 34  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 19  
 ; LENGTH: 38506  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 US-09-320-878-19



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2006, 13:11:49 ; Search time 1145 Seconds  
(without alignments)

10854.913 Million cell updates/sec

Title: US-10-618-976-1

Perfect score: 1503

Sequence: 1 atgcaactcaacgcgcttt.....ccgagccagccagcgct 1503

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1503	100.0	1503	9	US-10-618-976-1
2	287.4	19.1	1515	8	US-10-472-928-4401
3	287.4	19.1	5910	2	US-08-961-527-165
4	287.4	19.1	5910	7	US-10-158-844-165
5	287.4	19.1	2162598	8	US-10-472-928-4979
6	265.6	17.7	1377	9	US-10-617-320-336
7	159	10.6	1676	7	US-10-425-114-31307
8	134	8.9	1788	8	US-10-425-115-140665
9	104.2	6.9	540	7	US-10-767-701-19455
10	102.2	6.8	886	7	US-10-437-963-36082
11	99.2	6.6	2540	7	US-10-424-599-28136
12	84.2	5.6	2220	6	US-10-156-761-5375
13	84.2	5.6	9025608	6	US-10-156-761-1
14	76.8	5.1	563	7	US-10-424-599-28135
15	66.8	4.4	1945	7	US-10-425-114-29452
16	64.6	4.3	259	3	US-09-923-876-1014
17	64.6	4.3	259	3	US-09-923-876-1014
18	56.8	3.8	2217	3	US-09-938-842A-249
19	56.8	3.8	2217	3	US-09-938-842A-249
20	56.2	3.7	459	6	US-10-091-007-115
21	55.6	3.7	978	6	US-10-107-431-150
22	55.6	3.7	45055	6	US-10-107-431-277
23	52.8	3.5	3771	6	US-10-156-761-5283

24	50	3.3	36401	8	US-10-925-357-1	Sequence 1, Appli
25	49.4	3.3	1959	8	US-10-739-930-3025	Sequence 3025, Ap
26	48.6	3.2	745	6	US-10-336-597-7	Sequence 7, Appli
27	48.6	3.2	2241	6	US-10-156-761-1111	Sequence 1111, Ap
28	48.2	3.2	1146	6	US-10-163-198-12	Sequence 12, Appli
29	47.6	3.2	2658	7	US-10-282-122A-25660	Sequence 25660, A
C 30	47.4	3.2	1836	3	US-09-738-626-2483	Sequence 2483, Ap
C 31	47.4	3.2	1959	8	US-10-781-014-381	Sequence 381, App
C 32	47.4	3.2	2056	8	US-10-494-836-45	Sequence 45, Appli
C 33	47.4	3.2	2351	7	US-10-437-963-499	Sequence 499, App
C 34	47.4	3.2	3309400	3	US-09-738-626-1	Sequence 1, Appli
35	47.2	3.1	294	8	US-10-425-115-50318	Sequence 50318, A
36	47.2	3.1	1425	6	US-10-369-493-43028	Sequence 43028, A
37	47	3.1	1191	7	US-10-282-122A-14804	Sequence 14804, A
38	46.8	3.1	1191	7	US-10-425-114-24094	Sequence 24094, A
39	46.8	3.1	1123	8	US-10-425-115-103771	Sequence 103771, A
40	46.6	3.1	1545	6	US-10-163-198-96	Sequence 96, Appli
41	46.6	3.1	1947	9	US-10-928-992-25	Sequence 25, Appli
42	46.6	3.1	1948	7	US-10-260-238-298	Sequence 298, App
43	46.4	3.1	930	6	US-10-156-761-2723	Sequence 2723, Ap
44	46.4	3.1	1010	8	US-10-425-115-173269	Sequence 173269
45	46.4	3.1	1683	6	US-10-336-597-1	Sequence 1, Appli

#### ALIGNMENTS

#### RESULT 1

US-10-618-976-1  
; Sequence 1, Application US/10618976  
; Publication No. US20050074769A1  
; GENERAL INFORMATION:  
; APPLICANT: Fukuyama, Shiro  
; TITLE OF INVENTION: Polypeptides Having Glucanotransferase  
; FILE REFERENCE: 6012.200-US  
; CURRENT APPLICATION NUMBER: US/10/618,976  
; CURRENT FILING DATE: 2003-07-14  
; PRIOR APPLICATION NUMBER: US/09/687,360  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: PA 1999 01501  
; PRIOR FILING DATE: 1999-10-20  
; PRIOR APPLICATION NUMBER: PA 1999 01641  
; PRIOR FILING DATE: 1999-11-15  
; PRIOR APPLICATION NUMBER: 60/160,903  
; PRIOR FILING DATE: 1999-10-22  
; PRIOR APPLICATION NUMBER: 60/166,539  
; PRIOR FILING DATE: 1999-11-18  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1503  
; TYPE: DNA  
; ORGANISM: Thermus rubens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1503)  
US-10-618-976-1

Query Match	100.0%	Score 1503;	DB 9;	Length 1503;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1503;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGCAACTCAACGCCTTTGGAAATTTGCTCCACCCACCAAGTTTTCGGGTGCGTGG	60	
Db	1	ATGCAACTCAACGCCTTTGGAAATTTGCTCCACCCACCAAGTTTTCGGGTGCGTGG	60	
QY	61	GGGATTTGGGCTTGGCCGCGAGCGGTTTTCGACTGGCTGGCCGATGCGGA	120	
Db	61	GGGATTTGGGCTTGGCCGCGAGCGGTTTTCGACTGGCTGGCCGATGCGGA	120	
QY	121	GCCTGCTGGTGCAGTCTTTACCGCTGGGCCCTTACCAAGTTACCGGACTCGCGCTACCAAG	180	



Db	121	GCCCGCTGGTGCGCAGTCTTACCGCTCGGCGCCTACACGTTTACCGGCACTCGCGTACACG	180
Qy	181	TCCTTCTCGGCTTTTGC CGGTAAACCCGTA TTTGGTTGACCCCGAGATGCTGATTGA	240
Db	181	TCCTTCTCGGCTTTTGC CGGTAAACCCGTA TTTGGTTGACCCCGAGATGCTGATTGA	240
Qy	241	GGCTGGCTGGAA CAAAGCGAAGCGCCCGCGCTATCCGACCCAGCGCTGGA TTTATGCG	300
Db	241	GGCTGGCTGGAA CAAAGCGAAGCGCCCGCGCTATCCGACCCAGCGCTGGA TTTATGCG	300
Qy	301	TGGCTTTACCA GACCCCGCTCGCTCGCGCGGCTTTTCGCGGGTTTCGGGCAAGG	360
Db	301	TGGCTTTACCA GACCCCGCTCGCTCGCGCGGCTTTTCGCGGGTTTCGGGCAAGG	360
Qy	361	GCTTTCGGCCCA GGAATAAGACCGACTGGAAAGCCCTTTATCGAGCCGAGCGCTTCTGGCTG	420
Db	361	GCTTTCGGCCCA GGAATAAGACCGACTGGAAAGCCCTTTATCGAGCCGAGCGCTTCTGGCTG	420
Qy	421	GAAGACTATGCGCTCTTTATGGCCCTCA GACCCGGTTTGA CCGGTTTGA CCGGTAAGCCCTTGGAAACGAG	480
Db	421	GAAGACTATGCGCTCTTTATGGCCCTCA GACCCGGTTTGA CCGGTTTGA CCGGTAAGCCCTTGGAAACGAG	480
Qy	481	TGGAGCCCCGAGCTGCGCGACCGGTGAACCGGCTGCCCTGSCACAGGSCCGTGAGGAGCTG	540
Db	481	TGGAGCCCCGAGCTGCGCGACCGGTGAACCGGCTGCCCTGSCACAGGSCCGTGAGGAGCTG	540
Qy	541	GCCGAGAGGTGGCCCTTTTACGATGAGATTCA GATGGCTTTTATCTGGAATGGGCGCAG	600
Db	541	GCCGAGAGGTGGCCCTTTTACGATGAGATTCA GATGGCTTTTATCTGGAATGGGCGCAG	600
Qy	601	ACCAAGGCCCTATGCGGAATCCAAGGGGA TTTAGATTATCGCGGATATGCCCATTCTTTGTG	660
Db	601	ACCAAGGCCCTATGCGGAATCCAAGGGGA TTTAGATTATCGCGGATATGCCCATTCTTTGTG	660
Qy	661	GCTTTCGATTCTCAGATGTTCTGGGCCAA CCGCAGTACTTCTACCTCGAGGCGGATGCG	720
Db	661	GCTTTCGATTCTCAGATGTTCTGGGCCAA CCGCAGTACTTCTACCTCGAGGCGGATGCG	720
Qy	721	AACCCCAAGGTGGTGGGGGTTTCCGGGAGTACTTCTCCGAAACCGGCAAGCTCTG	780
Db	721	AACCCCAAGGTGGTGGGGGTTTCCGGGAGTACTTCTCCGAAACCGGCAAGCTCTG	780
Qy	781	GGCAATCGCTCTATCGTGGGATGTGATGSA AAGGACAAC TTTGCTGGTGCA TGGC	840
Db	781	GGCAATCGCTCTATCGTGGGATGTGATGSA AAGGACAAC TTTGCTGGTGCA TGGC	840
Qy	841	CGCATAGGCAAGTCTCAGCAGTGCCA CTTGTGTGCGATCGACCACTTCCCGGGGTTT	900
Db	841	CGCATAGGCAAGTCTCAGCAGTGCCA CTTGTGTGCGCATCGACCACTTCCCGGGGTTT	900
Qy	901	GAAGCCTACTGGGAGGTTCCGTTTGGCGGCCCAATGCTGTGAGGGGGCGCTGGGTCAA	960
Db	901	GAAGCCTACTGGGAGGTTCCGTTTGGCGGCCCAATGCTGTGAGGGGGCGCTGGGTCAA	960
Qy	961	GCCCCAGGGGAAGACTTTTGTCTCGGTGCGGCCCAACTGAGCGCATGCGCCCATATT	1020
Db	961	GCCCCAGGGGAAGACTTTTGTCTCGGTGCGGCCCAACTGAGCGCATGCGCCCATATT	1020
Qy	1021	GCCGAAGACTTGGGGGTGATCA CCCCAGAGGTGAGGCTTTGCGCGATGCTTCGGGTTG	1080
Db	1021	GCCGAAGACTTGGGGGTGATCA CCCCAGAGGTGAGGCTTTGCGCGATGCTTCGGGTTG	1080
Qy	1081	CCCGCATGAGATTTTGCAGTTTGTCTTTTCCGTTGAGGACAACGCTTTTTCGCCAC	1140
Db	1081	CCCGCATGAGATTTTGCAGTTTGTCTTTTCCGTTGAGGACAACGCTTTTTCGCCAC	1140
Qy	1141	AAC TACCCCGCGCA CGGCAATGTGTGTGTATACAGCGGAA CCCCAGCAACGACCA CACC	1200
Db	1141	AAC TACCCCGCGCA CGGCAATGTGTGTGTATACAGCGGAA CCCCAGCAACGACCA CACC	1200
Qy	1201	CTGGGATGTTTCCGCA CCGCGCCGAGGCCGAGCGGCTTTATGTCGGGCTACTTGGCC	1260
Db	1201	CTGGGATGTTTCCGCA CCGCGCCGAGGCCGAGCGGCTTTATGTCGGGCTACTTGGCC	1260

## RESULT 2

RES001 2  
US-10-472-928-4401

US-10-472-928-4401  
: Sequence 4401. Application US/10472928

Publication No. US20050020813A1

[illegible]

; GENERAL INFORMATION:  
: APPLICANT: CHIRON SpA

APPLICANT: CHIRON SPA

APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH  
TITLE OF INITIATION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS

; TITLE OF INVENTION: STREPTOCOCAL PHAGE. DO36026WO

FILE REFERENCE: P026926W0

; CURRENT APPLICATION NUMBER 200

;; CURRENT FILING DATE: 2003-09-26

; PRIOR APPLICATION NUMBER: GB-

; PRIOR FILING DATE: 2001-03-

; NUMBER OF SEQ ID NOS: 4979

```
; SOFTWARE: Seq
```

; SEQ ID NO 4401

; LENGTH: 1515

; TYPE: DNA

Query Match 19.14; Score 287.4; DB 8; Length 1515;

Query Match

Best Local Similarity	SI. %	Freq. NO.	Indels	Gaps
Matches 763. Conservative	0	Mismatches 696	18	Gaps 4

11 AACGGCGCTTTTGGGAATTATTGCTCCACCCCAACAGTATTTCCGGGTCCGCTGGGGGATTGGGG 70

[illegible]

130

QY 71 CTCTGGCCGCGAGGCCGAGCGGTTTGTGGACTGGCTGGCCGATGCGGGAGCCCGCTGGT 130

3

131 GGCAGGTCCTTACCGCTGGGCCCTACCAGTTACGGCGACTGCCCGTACCAGTCTTCTCTCGG 190

DB I28 GGCAATCCCTCCATTAGGAGCAGCATAGTACGGGGATTCCTAATCGTTTCTTCTTCT

191 CTTTGTGCGGTACCCGTAATTGGTTGACCCCGAGATGCTGATTGAAAAGGCTGGCTGG 250

Db 188 CCTTCGAGGAACACTCATTATTATCGATTATAGATATCTTGGTGGAGCAAGGTTGTGG 247

-----

Db  
248 AAGCAAGTGACCTTGAAGGAGTTGACTTTGGTAGCGATGCGTCTGAAGTTGACTATGCTA 307

[illegible]

308 A A T C T A T G C A C G T C G T C C T C T T T T A G A A A A G C G G T G A A C G T T C T T G A A G T C G 367

368 GAGATGTTAAAGATTTTTCAGAAAATTTGCTCAAGACAACCAATCATCGCTTCAGCTCTTTG 427  
QY 431 CGCTCTTTATGCGCCTCAAGACCCGGTTTTCAGCGCAAGCCCTGGAACGAGTGG--AGCC 487  
Db 428 CTGAGTATATGGCTATCAAGAGATATTTTGACAAATCTTGTGTGACTGAAATGGCCAGATG 487  
QY 488 CCGAGCTCGCGACCGGTGAACCGGCTGCCCTGGCCAGGGCCCGTGGAGAGCTGCCCGAGG 547  
Db 488 CAGATGCTCGTGTCTGTAAGCTTCAGCACITGAAAGCTATCGTGAGCAATGGCAGACA 547  
QY 548 AGTGGCCCTTTACGAGTGGAATTCAGTGGCTTTTATCTGGAATGGGGCCAGACCAAGG 607  
Db 548 AGTGGCTTTACCAACCGTGTGACTCAATACTTCTTCCAAACAATGGTTGAAATGGAAG 607  
QY 608 CCTATGCGGAATCAAGGGGATTCAGATTATCGCGGATATGCCATCTTTGTGGCTTCG 667  
Db 608 CTTAGCTTACGACACCAACATCGAATCGTTGGGACATGCCAATCTACGTAGCGGAAG 667  
QY 668 ATTCTCTCAGATGCTGGGCCAACCCGCAAGTACTTCTACCTCGAGCCGATGGCAACCCCA 727  
Db 668 ATTCAAGTGATATGTGGCAAAATCCACATCTCTTCAAAACAGATGTCAATGGTAAGCTA 727  
QY 728 CGGTGGTGGGGGGTTCGGCGGACACTTCTTCGAAAACCGGCCAGCTCTGGGGGCAATC 787  
Db 728 CTTGTATCGCAGGATGCCACAGATGAGTTTCTGTAACTGTGTCAGCTTTGGGGTAACT 787  
QY 788 CGCTCTATCGCTGGATGTGATGAAAGGGAACAATTTGCTGTGTGATTCGCCGCAATAA 847  
Db 788 CAATCTATGATCGGGAAGCAATGGAACAAGACGGCTACAATGTGTGATGAAACGCTTGC 847  
QY 848 GGCAGTGGCTCAAGCAGTGCACCTGGTGGCGCATCGACCACTTCGCGGGGTTTGAAGCCT 907  
Db 848 GTGAAAGCTTCAAAATCTACGATATCGTTGTATCGACCACTTCGTTGGCTTCGAATCT 907  
QY 908 ACTGGAGGTTTCCGTTTGGCGGCGCAATGCTGTGGAGGGCGCTGGGTCAAAGCCCGAG 967  
Db 908 ACTGGAAATCCCTGCTGGTTCGATACAGCAGCACCTGGTGTGAGTGGGTGAAAGSTCCAG 967  
QY 968 GGGAGAGCTGTTTGTGCTGGGTGGCGGCCCACTGAGCGATGCGCCCATCATTCGCCGAAG 1027  
Db 968 GTTACAAGCTTTTGTGACCGGTGAAAGAACTTGTGTGAGTAAACATATCCAGAAAG 1027  
QY 1028 ACCTGGGGGTGATCACCCCGAGTGGAGGCTTTGCGCGATGGCTTCGGGTTCGCCGGCA 1087  
Db 1028 ACCTGGCTTCATGACAGATGAGTATGATGATGCTGTAAGCTGACTGGCTTCCAGGAA 1087  
QY 1088 TGAAGATTTTGCAGTTTGTCTTTTCCGTTGAGGACAAGCGCTTTTGGCCCCCAACTACC 1147  
Db 1088 TGAAGATTTTCAATTTGCTTCAACCCAGAGAACGAAAGCATTTGATAGCCCACT--- 1144  
QY 1148 CCGGCAAGCAATGTGTGTGTATGACAGCGGAACCCAGCAACGACACCAACCTGGGAT 1207  
Db 1145 TGGCACCTGTCTAATCTAGTTATGTACAGGAACACACGATAAACATACGGCTTCTGGTT 1204  
QY 1208 GGTTCGCAACCGCGGAGCGGAGCGGCTTTCATGCGGGCTTCTGCGCGCTGATG 1267  
Db 1205 GGTACCGTATGATGATGATGATGCG-----ACTCGTGTATGATGCTGTTACA 1255  
QY 1268 GCATCGCTGTGTGTGCGGAATACAGAGTCGCGGCGCTTTGATGAGAGCTGGCTTCAAAA 1327  
Db 1256 CGAACCGTAAAGAATACGAACAGTGTG---ACACGCTATGCTTCGTACAGTATTTTCAT 1312  
QY 1328 GCCGGCCAGCTGGCTTATTTGTGCTTTGACGAGCTGCTGGGGCTGGGCGCCCGAGGCC 1387  
Db 1313 CAGTTAGCTTTATGGCAATTTGCAACTATGCAAGATTTTACTAGAAATGGATGAGCGAGCTC 1372  
QY 1388 GCATGAATTTCCCGGAGCGCTGGGGGACAACTGGGCGTGGCGCTACGCCGGAAGGCCACC 1447  
Db 1373 GTATGAATCTTCCCATCTACCTTGTGGTGAACACTGGTCTTGGCGTATGACTGAAGATCAAT 1432  
QY 1448 TCAGAGCCCGTCTGGCGCGGGAATGCGGGCCCTGGC 1484  
Db 1433 TGACACCAGCTGTGAGGAAGGTTTGTGCTGACTTGAC 1469

## RESULT 3

US-08-961-527-165

; Sequence 165, Application US/08961527

; Publication No. US20020323A1

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,527

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 165:

; SEQUENCE CHARACTERISTICS:

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; US-08-961-527-165

Query Match 19.1%; Score 287.4; DB 2; Length 5910;

Best Local Similarity 51.7%; Pred. No. 8.7e-70;

Matches 763; Conservative 0; Mismatches 696; Indels 18; Gaps 4;

QY 11 AACGCGCTTTTGGAAATTTTCTCCACCCACCAAGTTTTCGGGGTTCGCTGGGGGATTTGGGG 70

Db 108 AACGTCAAGTGTGTGTGTGATGACATCTTCTCTTCAGGAGCTTACGAATCGAT 167

QY 71 CTCTGGGCGCGAGGCGGAGCGGTTTTTGGACTGGCTGGCCGATGCGGGAGCCCGCTGGT 130

Db 168 CATTGGTCAAAGTGTCTACGACTTGGTTGATTTCTTGGTCCGTACAAAACAACGTTACT 227

QY 131 GGCAAGTCTTACCGTGGGCGCTTACCGGACTTACCGGACTCGCGTACAGTCTTCTTCGG 190

Db 228 GGCAATCTCTTCATTAGGAGCAACTAGTTACGGGATTTCTCTTACCAATCTTCTCAG 287

QY 191 CTTTTCGGGTAAACCGGTATTTGGTTGACCCCGAGATGCTGATTTGAAAAAGGCTGGCTGG 250

Db 288 CCTTCGAGGAACACATCAATTTATCGATTTAGATATCTTGGTGAGCAAGGTTGTTGG 347

QY 251 AACAAAGGAGAGGCGCCCGCGCTATCCGAGCCAGCGCGTGGATTTATGGCTGGCTTACC 310

Db 348 AAGCAAGTGACCTTGAAGGAGTTGACTTTTGGTAGCGATGCGTCTGGAAGTTGACTATGCTA 407

QY 311 AGACCCGCTGGCCCTGTTTGGCGGGGCTTTTCGGGCAAGGCTTCGGGCC 370

Db 408 AAATCTACTATGACGCTGCTCTCTTTTAGAAAAAGCGGTGAACGTTTCTTTGAAGTCG 467

QY 371 AGGATTAAGACCCGACTGGAAGCCTTTATCGAGGCCGAGCGCTTCTGGCTGGAAGACTATG 430

Db 1533 TGACACACGCTGTCGAGGAAGGTTTGTCTGACTTGAC 1569

RESULT 4

US-10-158-844-165

Sequence 165, Application US/10158844

Publication No. US20040029118A1

GENERAL INFORMATION:

APPLICANT: Kunsch et al.

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-R

COMPUTER: Dell Latitude Pentium 3

OPERATING SYSTEM: Windows 98

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/158,844

FILING DATE: 03-Jun-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/961,527

FILING DATE: 1997-10-30

APPLICATION NUMBER: US 60/029,960

FILING DATE: 1996-10-31

ATTORNEY/AGENT INFORMATION:

NAME: Hyman, Mark J.

REGISTRATION NUMBER: 46,789

REFERENCE/DOCKET NUMBER: PB340P1D1

INFORMATION FOR SEQ ID NO: 165:

SEQUENCE CHARACTERISTICS:

LENGTH: 5910 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 165:

US-10-158-844-165

Query Match 19.1%; Score 287.4; DB 7; Length 5910;

Best Local Similarity 51.7%; Pred. No. 8.7e-70;

Matches 763; Conservative 0; Mismatches 696; Indels 18; Gaps 4;

Qy 11 AACGCGCTTTTGGAAATTTTGTCCACCCACCCAGTTTTCGGGTCGCTGGGGGATTTGGG 70

Db 108 AACGTCAAAGTGGTGTGATGCAATCTCTTCTTCAGGAGCTTACGGAATCGAT 167

Qy 71 CTCTGGGCGCGAGCGCGGATTTTGGACTGGCTGGCGGATGCGGGAGCCCGCTGGT 130

Db 168 CATTTGGTCAAGTGTACGACTTGTGATTTCTTGTTCGATCAAAACAACGTTACT 227

Qy 131 GGCAGGCTTTACCGCTGGGCGCTTACGAGTTACGCGGCTGCGCGTACAGTCTTCTCGG 190

Db 228 GGCAGGCTTTTCCATTAGGAGCAACTAGTTACGGGATTTCTTCTTACCAATCTTTCTCAG 287

Qy 191 CTTTGGCGGTAAACCGTATTTGGTTGACCCCGAGATGCTGATTAAGAAAGCTGGCTGG 250

Db 288 CTTTCGAGGAAACACTCATTTTATCGATTAGATATCTTGTGAGCAAGTTGTTGG 347

Qy 251 AACAAAGCAAGCGCCCGCTATCCGACCCGCGCTGATTCGACCCGCGCTGATTCGCTGGCTTTACC 310

Db 348 AAGCAAGTACCTTGAAGGAGTTGACTTTTGTAGCGATGCGTCTCAAGTTGACTATGCTA 407

Qy 311 AGACCGCTGGCCCTGTGTGGCGGGCTTTCGGGGGTTTCGGGCAAGGCTTCGGCCC 370

Db 408 AAATCTACTATGACGCTGCTCTCTTTAGAAAAAGCGGTGAAACGTTTCTTTGAAGTCG 467

Db 468 GAGATGTTAAAGATTTTGAGAAATTTGCTCAAGACAACCAATCATGGCTTGAGCTTTTG 527

Qy 431 CGCTCTTTATGGCCCTCAAGACCCGGTTTGAAGCAAGCCCTGGAACGAGTGG---AGCC 487

Db 528 CTGAGTATAGGCTATCAAGAGTATTTTGACATCTTGTCTGACTGATGGCCAGATG 587

Qy 488 CCGAGCTCGCGGACCGGTGAACCCGGCTGGCCAGGCGCGGTGAGAGCTGCGCCAGG 547

Db 588 CAGATGCTCGTCTCGTAAAGCTTCAGCACTTGAAGCTATCGTGAGCAATTTGCGAGCA 647

Qy 548 AGTGGCCCTTTACGAGTGGATTGAGTGGCTTTTATCTGGAATGGCGGCGAGCAAGG 607

Db 648 AGTTGGTTTACCACCGGTGACTCAATCTCTTCTTCCAAATGTTGTAATTTGAAG 707

Qy 608 CTTATGCGGAATCAAGGGGATTCAGATTATCGCGGATATCCCATCTTTTGTGGCTTTG 667

Db 708 CTTACGCTTAACGACACCAACATCGAATCGTTGGGACATGCCAATCTACGTAGCGGA 767

Qy 668 ATTCTCAGATGTCTGGCCCAACCGCAGTACTTCTACTCTGAGCGGATGGCAACCCCA 727

Db 768 ATTCAAGTATATGTGGCAAAATCCACATCTCTTCAAAACAGATGTCAATGGTAAGGCTA 827

Qy 728 CGGTGGTGGCGGGCTTCCGGGAGTACTTCTCGAACCAGCGCTCTGGGGCAATC 787

Db 828 CTTGTATCGCAGGATGCCCCAGATGATGTTTCTGTAACTGGTCACTTTGGGGTAACT 887

Qy 788 CGCTCTATCGTGGGATGTGATGAAAGGACAACTTTGCTGTGTGATTCGCCGCATAA 847

Db 888 CAATCTATGACTGGAGCAATGNAAGACGGCTACAAATGTTGGATTGAACGCTTGC 947

Qy 848 GGCAGTCTCCTCAAGCAGTCCACTGTGTGCGATCGACCACTTTCGCGGGTTTGAAGCT 907

Db 948 GTGAAAGCTTCAAAATCTACGATATCGTTCTGTATCGACCACTTCGCTGGCTTCAATCT 1007

Qy 908 ACTGGAGGTTCCGTTTGGCGGCCAATCTGTGGAGGGCGCTGGGTCAAGCCCGAG 967

Db 1008 ACTGGGAATCCCTGCTGTTCCGATACAGCAGACCTGGTGGTGAAGGTCCAG 1067

Qy 968 GGGAGAGCTGTTGCTCGGTGCGGCGCCAACTGAGCGATGCGGCCATCATTTGCCGAAG 1027

Db 1068 GTTACAGCTTTTGCACCGCTTAAAGGAAGACTTGTGTGAGTAAACATCATCGCAGA 1127

Qy 1028 ACCTGGGGTATCAACCCCGAGTGGAGGCTTTGCGGATGCTGCTGGGTTCCCGGCA 1087

Db 1128 ACCTGGCTTCATGACAGATGAAGTATCGAATTCGTTGAACGTAAGTCTGGCTTCCAGGA 1187

Qy 1088 TGAAGATTTGAGTTGCTTTTCCGTGAGGACAGCCCTTTTGGCCCGACACTACC 1147

Db 1188 TGAAGATTTCTCAATTTGCTTCAACCCAGAGACGAAAGCAATGATAGCCCACT--- 1244

Qy 1148 CCGGCAAGCAATGTTGTTGTGTATACGCGAACCACGACAAACACACACCTTGGGAT 1207

Db 1245 TGGACCTGCTTAATCTAGTATGATACAGGACACACGATTAACATAGGTTCTTGGTT 1304

Qy 1208 GTTTCGCGACCGCGCGGAGCGGCGCTTCAAGCGGCTTACCTGGCGCGCTATG 1267

Db 1305 GGTACCGTAATGAGATTGATGTCG-----ACTCGTGAATACATGCTGCTTACA 1355

Qy 1268 GCATCCGTTTGTTCGGAATACGAGTTCGCGGCGCTTTGATCAGCTGGCTTCAAA 1327

Db 1356 GCAACCGTAAAGAAATACGAAACAGTGT---ACACGCTATGCTTCGTACAGATTTCAT 1412

Qy 1328 GCCCGGCAAGCTGCTATTTGCTTTCAGGACGTTGCGGGCTGGCGCCCGAGGCC 1387

Db 1413 CAGTTAGCTTTATGGCAATTGCACTATGCAAGATTACTAGAAATGATGAGGAGCTC 1472

Qy 1388 GCATGAATCTTCCCGGACGGCTGGGGACAACTGGCGGTGGCGCTTACGCGGAGCGGACC 1447

Db 1473 GTATGAATCTTCCATCACTCCCTTGTGGAAACTGTGCTTGGGGTATGACTGAAGATCAAT 1532

Qy 1448 TCGAGCCCGGTTCGCGCGGAGCTGCGGGCCCTGGC 1484



```
Qy 668 ATTCTCAGATGCTCTGGGCAACCGCAGTACTTCTACCTCGAGCGCGATGGCAACCCCA 727
Db 1913201 ATTCAAGTGATATGTGGGCAATCACATCTCTTCAAAACAGATGTCAATGGTAAGGCTA 1913142
Qy 728 CGGTGGTGGCGGGGCTTCCGGGGGACTACTTCTCGAAACCGCGCAGCTCTGGGGGCAATC 787
Db 1913141 CTTGTATCGCAGGATGCCACAGATGAGTTTCTGTAACTGGTCAGCTTTGGGGTAATC 1913082
Qy 788 CGCTCTATCGTGGGATGTATGGAAGGACAACTTTGCTGGTGCAATGCCCGCATAA 847
Db 1913081 CAATCTATGACTGGGAAGCAATGGACAAGACGGCTACAAATGGTGGATGAACGCTTGC 1913022
Qy 848 GGCAGTCCTCAAGCAGTGCACCTGTGCGCATGACCACTTCGCGGGTTTGAAGCT 907
Db 1913021 GTGAAGCTTCAAAATCTACGATATCGTTCGTATCGACCACTTCGCTGGCTTCAATCTT 1912962
Qy 908 ACTGGAGGTTTCCGTTTGGCGGCGCAATGCTGTGGAGGGCGCTGGGTCAAGCCCCCAG 967
Db 1912961 ACTGGGAATCCCTGCTGTTCCGNATCAGCAGCACCTGGTGAGTGGTGAAGGTCCAG 1912902
Qy 968 GGGAGAAGCTGTTTCTCGCGTGGCGGCCCAACTGAGCGATGCGGCCCATCATATGCCGAAG 1027
Db 1912901 GTTACAAGCTTTTGCAGCGGTTAAGGAAGAACTTGGTGAGCTAAACATCATCGCAGAAG 1912842
Qy 1028 ACCTGGGGGTGATCACCCCGAGGTGGAGGCTTTGGCGGATGCGCTTCGGGTTCCCGGCA 1087
Db 1912841 ACCTTGGCTTCATGACAGATGAAGTATCGAATTCGCGTGAACGTACTGGCTTCCCGGAA 1912782
Qy 1088 TGAAGATTTTTCAGTTTCTTTTCCGCTGAGGACAGCGCTTTTGGCCCACTACCTACC 1147
Db 1912781 TGAAGATTTCTCAATTTGCTTCAACCGAAGAGCAAGCAATTCATAGCCCACT--- 1912725
Qy 1148 CCGGCGCAGGCAATGTGTGTGTGTACAGCGGAACCCAGCAACAGACACCACTGGGAT 1207
Db 1912724 TGGCACCTGTCTAATTTGCTTCAACCGAAGAGCAAGCAATTCATAGCCCACT--- 1912665
Qy 1208 GGTTCGGCACCGCGCGGAGGCGGCGGCTTTCATCGGCGCTTACCTGGCGGCTATG 1267
Db 1912664 GGTACCGTAATGAGATTGATATGCG-----ACTCGTGAATGATGCTCGTTACA 1912614
Qy 1268 GCATCCGTTGTTTGTGGAATACGAGTTCGGGGGCTTTCATGAGCTGGCTTCAAAA 1327
Db 1912613 GCAACCGTAAAGATACGAACAGTGT---ACAGCTATGCTTCGTACATATTTTCAT 1912557
Qy 1328 CCGCGGCAAGCTGTGCTATTTGCTTTCAGGACGCTGCTGGGCTGGGCGGCGGCGCC 1387
Db 1912556 CAGTTAGCTTTATGGCAATTCGAATATGCAAGATTTACTAGAAATTGGATGAGGCGCTC 1912497
Qy 1388 GCATGAACCTTCCCGGAGCTGGGGGCAACTGGCGCTGGGCTTACCGCAAGCGGACC 1447
Db 1912496 GTATGAACCTTCCCATCTACCTTGGTGGAACCTGGTCTTGGCGGTATGACTGAAGATCAAT 1912437
Qy 1448 TCGAGCCGCTGCTGCGCGGAGCTTGGCGGCTGGC 1484
Db 1912436 TGACACCACTGTCGAGGAAGTTTGTGACTTGAC 1912400
```

## RESULT 6

```
US-10-617-320-336
; Sequence 336, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNO
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
```

```
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/617,320
FILING DATE: 10-Jul-2003
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
APPLICATION NUMBER: 60/ 085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 336:
SEQUENCE CHARACTERISTICS:
LENGTH: 1377 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1377
SEQUENCE DESCRIPTION: SEQ ID NO: 336:
US-10-617-320-336
```

```
Query Match 17.7%; Score 265.6; DB 9; Length 1377;
Best Local Similarity 52.1%; Pred. No. 8.5e-64;
Matches 698; Conservative 0; Mismatches 624; Indels 18; Gaps 4;

Qy 148 GGCCTACCACTAGTACGGGAGTCCCGTACAGTCCCTCTCGGCTTTTGGCGGTAACCCG 207
Db 4 GGAGCAACTAGTTACGGGATTCCTTACCAATCTTTCTCAGCCTTCGCGAGGAACACT 63
Qy 208 TATTTGGTTGACCCCGAGATGCTGATTGAAAAAGGCTGGTGGAAACAAAGCGAGCGCC 267
Db 64 CATTTTATCGATTAGATATCTTGGTGAGCAAGTTTGTGGAAGCAAGTGAACCTTGA 123
Qy 268 CCGCGGTATCCGACCCAGCGGTGATATATGGCTTTTACCAAGCCCGCTGGCCCTG 327
Db 124 GGAGTTGACTTTGGTAGCGATGCTCTGAAGTTGACTATGTAAATCTACTATGACGT 183
Qy 328 TTGCGGGGGCTTTCGGGGGTTTCGGCCAGAGGCTTCGGCCAGGATAAGACCCGACTG 387
Db 184 CGTCTCTTTTGAAGAAAGCGGTGAAACGTTCTTTTGAAGTCGGAGATGTTAAAGATTT 243
Qy 388 GAAGCCTTTATCGAGGCGGAGCGCTTCTGGCTGGAAGACTATGGCTCTTTATGGCCCTC 447
Db 244 GAGAAATTTGCTCAAGACAACCAATCATGGCTTGAGCTCTTTTGTGAGTATATGCTATC 303
Qy 448 AAGACCCGGTTTGAAGCAAGCCCTGGAACAGAGTGGAGC---CCGAGCTCGGAGCGT 504
Db 304 AAGAGCAATTTTGACAATCTTCTTGGACTGAATGGCCAGATGAGATGCTCGTGTCTGT 363
Qy 505 GAACCGGCTGCCCTGGCCAGGCGGCTGAGAGCTGGCCGAGGAGTGGCCCTTTACGAG 564
Db 364 AAGCTTTCAGCACTTGAAGCTATCTGAGCAATTTGGCAGCAAGTTGGTTTACACCGT 423
Qy 565 TGGATTGAGTGGCTTTTATCTGGAATGGGGCCAGACCAAGGCTTATGCCGAATCCAAG 624
```

424 GTGACTCAATCTCTCTTCCACATGTTGAAATTTGAAGCTTACGCTACAGCAAC 483  
425 GGAATTCAGATATTCGGCGATATGCCCATCTTTGTGGCTTTCGATTCCTCAGATGCTGG 684  
484 CACATCGAAATCGTTGGGACATGCCAATCTACGTAGCGAAGATTCAGATGATATGG 543  
685 GCCAACCCGAGTACTTCTACCTCGAGCGCGATGGCAACCCACGGTGGTGGCGGGTT 744  
544 GCAATCCCATCTCTTCAAAACAGATGTCAATGGTAAGGCTACTTGTATCGCAGGATGC 603  
745 CCGCGGAGTACTTCTCCGAAACCGCCAGCTCTGGGGCAATCGCTCTATCTCGCTGGAT 804  
604 CCACCCAGATGATTTCTGTAACTGGTCAGCTTTGGGGTATCCATCTATGACTGGAA 663  
805 GTGATGAAAGGACAACTTTGCTGTGTCGATGATCCCGCATAAAGCGATCGCTCAAGCAG 864  
664 GCAATGGACAAAGACGGCTACAAATGGTGGATGAACGCTTGGGTGAAGCTTCAAAATC 723  
865 TGCACCTGTGTGGCATCGACCACTTCCGGGGTTTGAAGCCTACTGGGAGGTTCCGTTT 924  
724 TACGATATCGTTGATTCGACCACTTCCGGTGGCTTGGATCTTACTGGGAAATCCCTGCT 783  
925 GCGCGGCCCAATGCTGTGGAGGGCGCTGGGTCAAGCCCGGAGGAGCTGTGTTGCT 984  
784 GGTTCGATACAGACGACCTGTGTAGTGGGTGAAGGTCAGGCTACAGCTTTTGA 843  
985 GCGGTGCGGCCCACTGACGATGCGCCCATCATGTCGGAAGACCTGGGGGTGATCAC 1044  
844 GCGTTAAGGAAGAACTTGGTGGCTTAAACATCATCGCAGACACCTTGGCTTCATGACA 903  
1045 CCGAGGTGGAGCTTTGGCGATGCTTCCGGCTTCCCGGCGATGAGATTTTGCAGTTT 1104  
904 GATGAAGTGAATGAAATGGCGTGAACGTAAGTGGCTTCCAGGAATGAAGATTTCTCAATT 963  
1105 GCTTTTTCGGGTGAGGACACGCTTTTTCGCCCACTTACCCCGGCGCAGCAATGTG 1164  
964 GCCTTCAACCCAGAGACGAAAGATTTGATAGCCCACT--TGGCACCTGTACTCA 1020  
1165 GTGGTGTACAGCGGAAACCCACGACAAACAGACACCACTGGGATGGTTCCGACCGCGCG 1224  
1021 GTTATGTACAGAGAAACACAGGATAACATACGCTTCTGGTGGTACCGTAATGAGATT 1080  
1225 GAGCCGAGGGGCTTCTACGCGGCTTACCTGGCGGCTATGGCATCGTTGTTGTCG 1284  
1081 GATGATGCG-----ACTCGTGAATACATGGCTGTTTACAGAACCGTAAAGAAATC 1131  
1285 GAATACAGGTGCGGGCGCTTTGATCGAGCTGGCTTCAAAAGCCCGCAGCTGGCT 1344  
1132 GAAACAGTGT---ACACGCTATGCTTCGTACAGTATTTTCATCAGTTAGCTTTATGGCA 1188  
1345 ATTGTGCTTTTGACGAGCTGTGGGCTGGGCGCCGAGCCCGCATGAACCTTCCCGGA 1404  
1189 ATTGCAACTATGCAAGATTTTACTAGAATTGATGAGGACGCTCGTATGAATTTCCCATCT 1248  
1405 CGGCTGGGGAACAACTGGGCGTGGCGTACCGCGAAGGCGACCTCGAGCCCGCTGTGGCC 1464  
1249 ACCCTTGGTGAACACTGGCTTGGCGGTATGACTGAAGATCAATTGACACACGAGTGTGAG 1308  
1465 GCGGAGTGGGGCCCTGGC 1484  
1309 GAAGGTTTGTGCTTACITGAC 1328

## RESULT 7

US-10-425-114-31307  
; Sequence 31307, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E

APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 31307  
LENGTH: 1676  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURES:  
OTHER INFORMATION: Clone ID: UC-ZMFLB73152G04\_FLI  
US-10-425-114-31307

Query Match 10.6%; Score 159; DB 7; Length 1676;  
Best Local Similarity 50.5%; Pred. No. 5,4e-34;  
Matches 413; Conservative 0; Mismatches 402; Indels 3; Gaps 1;

415 TGGCTGGAAGACTATGCGCTCTTTATGGCCCTCAAGACCGGTTTGACGGCAAGCCCTGG 474  
379 TGGCTTGAAGATGACGCTCTTTTGTCTGCTATTGACAAAAGCATCAATGCAATCTCTTGG 438  
475 AACGAGTGGAGCCCGGAGCTGGCGGACCGTGAACCGGCTGCCCTGGCCAGGGCCGTGAG 534  
439 TACGAGTGGCTGAACCTCTGAAAATCGCCATCTTGGAGCCTTGGAGATATATATGAA 498  
535 GAGCTGGCGGAGGAGGTGGCCCTTTACGAGTGAATTCAGTGGCTTTTATCTGGAATGG 594  
499 AAACAGAGGATTTTATAGAGACATTTATGGCTCAGCAGTTCTTATTTCAAGGCAATGG 558  
595 GGCAGACCAAGCCCTATCCGAAATCCAAAGGGATTCAGATTCATCGCGATATGCCCATC 654  
559 CAGCGGATTCGTAATATGACAAAAGCTGGGTATCAGCATCATGGTGAATGCTCTATA 618  
655 TTTGTGGCTTCGATTCCTCAGATGCTGGGCGCAACCGCAGTACTTCTACCTCGAGGCC 714  
619 TATGTTGGCTACCATAGCGCAGATGTTGGCGCAACAGGAATCAATTTTGTGGACAA 678  
715 GATGCAACCCCAACGGTGGTGGCGGCTTCCGCGGAGTACTTCTCCGAAACCGGCGAG 774  
679 AACGGTTTCCCACTTTCGTTAGTGGCTTCCACCTGATGATTCATTTAGTGAACCGGCTCA 738  
775 CTCTGGGCAATCCGCTCTATCGCTGGATGATGAGAGGAGCAACTTTTGTCTGTGTC 834  
739 CTATGGAACAGTCCATTTGACGACTGGAAGCTATGGAAGCAATGTTTTTTCATGTGG 798  
835 ATTGCGCGCATTAAGGAGTCCGCTCAAGCAGTGCACCTGGTGGCATCGACCACTTCCGC 894  
799 ATAAGAGAGATTAAGCTGCCCTTGTATGATGATGATTCGATTTGACCATTTCCGG 858  
895 GGGTTTGAAGCCCTACTGGAGGTTCCGTTGGCGGCCCAATGCTGTGAGGGGCGCTGG 954  
859 GGGCTTGGGGTTTGGGCGAGTCCCTTCTGTGACAAAAGTAGCACTGGTGGAGCTGG 918  
955 GTCAAAGCCCAAGGAGAGAGCTGTTGCTGGTGGCGGCCCACTGAGCGATGCGCCC 1014  
919 AGGGCTGGACCAAGGAATAGCTTTTGTGACAGCTCTTCAAGCTGTGTGTAAGATAGAT 978  
1015 ATCATTTGCCGAAGACCTGGGGGTGATCACCCCGAGGTGGAGGCTTTGCCGATGGCTTC 1074  
979 ATAATAGCAAGAGATCTGGGGGTAAATCTGAAGATGTGTTTCAGCTAAGGAATCAATT 1038  
1075 GGGTTCCCGGATGAAGATTTTGCAGTTTGTCTTTTCCGGTGAAGCAACAGCTTTTGTG 1134  
1039 GGGGCGCTGGGATGGCAGTTCTCCAGTTTGTCTTCCGAGGTGGTCTTGCACAAACCTCAT 1098  
1135 CCCCACAACTACCCCGGCAACGCAATGTTGTTGTGATGAGCGGAACCCACGACAGCAC 1194  
1099 TTGCCACACACCATG---AAATGGATCAAGTTGTGTACACTGGAACACATGATACGAT 1155  
1195 ACCACCTGGGATGGTTTCGCGACCGCGCGAGGCGGA 1232



Db 1156 ACAGTCTTGGCTGGTGGCAAAATTTACGAGGAAGA 1193  
|||||

## RESULT 8

US-10-425-115-140665  
; Sequence 140665, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 140665  
; LENGTH: 1788  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_59770C.1  
US-10-425-115-140665

Query Match 8.9%; Score 134; DB 8; Length 1788;

Best Local Similarity 52.4%; Pred. No. 5.3e-27;  
Matches 319; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

Qy	624	GGGGAATTCAGATTATCGCGGATATGCCCATCTTTTGGGCTTCGATTCCTCAGATGCTG 683
Db	586	GCCTATCAGCATCATGGGTGACATGCTCTATATGCTGTGCTACCATAGCGCATATGTTG 645
Qy	684	GGCCAAACCGCGACTCTTCTACCTCGAGCGGATGCGAACCCCAACCGTGGTGGGGCGT 743
Db	646	GGCGAAGCAGGAATCATTTTTCCTGGACAAAACCGTTTCCCACTTTTCGTTAGTGGCGT 705
Qy	744	TCCCGCGGACTACTCTCCGAAACCGGCGACTCTGGGCGCAATCCGCTCTATCGCTGGGA 803
Db	706	TCCACCTGTATGATTTAGTGAAACGGGTCACTATGGACAGTCCATTTGACGACTGAA 765
Qy	804	TGCGATGAAAGGACAACTTTTGGCTGTGTCATTTGCCCGCATTAAGGAGTGCCTCAAGCA 863
Db	766	AGCTATGGAAGCAGATTGTTTTCATGTTGATTAAGAGGATTAACCGTGCCTTGATTT 825
Qy	864	GTGCCACTGTGGCATCGACCACTTCCGGGGTGTGAACCTACTGGGAGGTTCGGTT 923
Db	826	GTATGATGAATTCGGTATTTGACCACTTCCGGGGGCTTGGGGTTTTGGGAGTCCCTTC 885
Qy	924	TGGCGGGCCCAATGCTGTGGAGGGCGCTGGGTCAAGCCCGAGGGAGAGCTGTTTGC 983
Db	886	TGATGCAAAAGTAGCATGTTTGGTGGAGCTGAGGGCTGGACCAAGGATAGCTTTTGA 945
Qy	984	TGCGGTGGGGCCCAACTAGCGATGCGCCCATCATTTCCGGAAGACCTTGGGGGTATCAC 1043
Db	946	CACGCTCTTCAAGCTGTGTGTAAGATAGATATAATAGCAGAAGATCTGGGGGTAAATTAC 1005
Qy	1044	CCCCGAGGTGAGGCTTTGCGCGATGGCTTCGGGTTCCTCCGCGCATGAGATTTTGCAGTT 1103
Db	1006	TGAAGATGTCTTCAGCTAAGGAAATCATTTGGGGCCCTTGGGATGGCAGTTCTCCAGTT 1065
Qy	1104	TGCTTTTTCGGTGGAGCAACGCTTTTTCGCCCACTACCCCGCGCAGCGCAATGT 1163
Db	1066	TGCTTTTCGAGGTGTTCTGCAACCTCATTTTGCACACACCATG---AAATGGATCA 1122
Qy	1164	GGTGGTGTACAGCGGAACCCACGACCAACCACTCTGGGATGTTCCGCGCGCGCC 1223
Db	1123	AGTTGTGTACACTGGAACACATGATTAACGATACAGTCTTGGCTGGTGGCAAAATTTACC 1182
Qy	1224	GGAGGCCGA 1232 

Db 1183 AGAGGAAGA 1191

## RESULT 9

US-10-767-701-19455  
; Sequence 19455, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 19455  
; LENGTH: 540  
; TYPE: DNA  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3480-052-P1-K1-G9  
US-10-767-701-19455

Query Match 6.9%; Score 104.2; DB 7; Length 540;

Best Local Similarity 52.6%; Pred. No. 9.3e-19;  
Matches 251; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

Qy	757	TTCTCGAAACCGGCCAGCTCTGGGCAATCCGCTCTATCGCTGGGATGTGATGGAAGG 816
Db	6	TTTAGCGAAACGGGTCAACTATGGAACAGTCCATTGTACGACTGGAAGCTATGGAAGCA 65
Qy	817	GACAACTTTCCTGGTGCATTTGCCGCAATAGGCAATGCTCTCAGCAGTGCACCTGGTG 876
Db	66	GATGGCTTTTTCATGGTGGACAAAGAGGATTAACGTCCTTGATTTGATGATGAGTTC 125
Qy	877	CGCATCGACCACTTCCGGGGTTTGAAGCCTACTGGGAGGTTCGGTTTGGCGGCCCAAT 936
Db	126	CGTATTGACCATTTTCCGGGGTCTTGAGGTTTTTGGCAGTCCCTTCTGATGCAAAAGTA 185
Qy	937	GCTGTGGAGGGGGCTGGGTCAAGGCCCGAGGGGAGAGCTGTTTGTGCGGTGCGGGCC 996
Db	186	GCATGCTGGTGGAGCTGGAGGGCTGGACCAAGGAATAGCTTTTGTGACGGCTCTTCAAA 245
Qy	997	CAACTGAGCGATGGCCCATCATTTGCCGAAGACTGCGGGGTGATCACCCCGAGGTGAG 1056
Db	246	GCTGTTGGTGAATAAATAATAATAGCAGAAGACCTGGGGGTAAATTACTGAAGATGTTT 305
Qy	1057	GCTTTGGCGGATGGCTTGGGTTTCCCGGCATGAAGATTTTGCAGTTTGTCTTTTCCGGT 1116
Db	306	CAGTAAGGAATTCATTTGGGGCCCTTGGGATGGCAGTTCTCCAGTTTGTCTTTGGAGT 365
Qy	1117	GAGGACAAACCCCTTTTTCGCCCACTACCCCGCAGCAATGTTGGTGGTGTACAGC 1176
Db	366	GGTTCTGCAACCTCACTTGCCTTGCCTTGGGATGGCAGTTCTCCAGTTTGTCTTTGGAGT 422
Qy	1177	GGAAACCCACGACAGCAGCACCACTCGGATGTTTCCGACCGCGCGGCGGCGAG 1233
Db	423	GGAACATGATAACGATACAGTCTTGGTGGTGGCAAAATTTTCCAGAGGAAGAG 479

## RESULT 10

US-10-437-963-36082  
; Sequence 36082, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.



FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 5375  
LENGTH: 2220  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(2220)  
US-10-156-761-5375

Query Match 5.6%; Score 84.2; DB 6; Length 2220;  
Best Local Similarity 46.6%; Pred. No. 4.5e-13;  
Matches 389; Conservative 0; Mismatches 428; Indels 18; Gaps 3;  
391 GCCTTTATCGAGCGCGAGCGCTTCTGGCTGGAAGACTATGCGCTCTTTATGGCCCTCAAG 450  
Db GACTTCTCGCGAGAGGGCGAGCGCTGAGGACCAACGCGAGCTGTGTGGCGCT---G 1023  
451 ACCCGGTTTGACGCAAGCCCTGGAACGAGTGGAGCCCGAGCTGCGGACCGTGAACCG 510  
Db GCCGAGGTGTACGGGGCGGATGGGACGCTGGCGCGCGCGCTGCGGACCGCGCTCG 1083  
511 GCTGCCCTGCGGAGCGCGCTGAGGAGTGGCGGAGGTGGCCCTTACGATGGATT 570  
Db GCGGAGACCGCGCGCGCGCGCGAGTGTGATGACCGCGTCTCCACGCTGGCTC 1143  
571 CAGTGGCTTTTATCTGGAATGGGGCCAGACCAAGGCTTATGCGAATCAAGGGGATT 630  
Db GCCTGGCTGACGAGACAGCTGCCCGCGCTCAGCGGGCGCGCTGAGCGGGCATG 1203  
631 CAGATTATCGGCGATATGCCCATCTTTGTGGCTTTCGATCTCTAGATGTCTGGGCCAAC 690  
Db GCGGTGCGACTCGTGACGACCTCGCGGTGCGCGTCCATCCGTCGCGCGCGACGCGCTGG 1263  
691 CCGGAGTACTTCTACCTGAGCGCGATGGCAACCCAGCGTGGTGGCGGGGTTCCGGGG 750  
Db GCGGAGGAGGATCTTCCGCGCGGCA-----TGTCGGTGGGGCGCGCGCGCC 1311  
751 GACTACTTCTCGAAACCGCGAGCTTGGGGCAATCCGCTCTATCGCTGGGATGTGATG 810  
Db GACGCTTTCAAGCGCTCGGACGAGCTGGGGCTTGGCGCTTGGCGACCGGACCGCTC 1371  
811 GAAAGGCAAACTTTGCTGTGCAATGCCCGCATAGGCGAGTCTCTCAAGCAGTGCAC 870  
Db GCCGAGTCCGGTACGCGCGCTACGCGCGTCTGCTCCGCGCGCTCTCCGATACGCGGGC 1431  
871 CTGGTGGCATGACCACTTCGCGGGTTTGAAGCTTACTGGAGGTTCCGTTGGCGCG 930  
Db GCCCTGCGCATGACCACTGATGGGGCTTTCGCGCTGTGGTGGTGGCGAGGGCGG 1491  
931 CCCAATGTGTGGAGGGCGCTGGGTCAAAGCCCCAGGGGAGAACT---GTTTCTGCG 987  
Db CCGCCACGAGGGAGCACTGCTCGGTACGACGCGGAGGCGATGCTCGCGATCTCTGTC 1551  
988 GTGCGGGCCCACTGAGCGATGCGCCCATCATTTGCGAAGACTTGGGGGTGATCAACCCC 1047  
Db CTGGAGGATCGCGACCGGGGCACTGGTGTATCGGAGGAGCTCGGACCGTCAACCC 1611  
1048 GAGTGGAGGCTTTGCGGATGGCTTCGGTTCGCGGATGAAGATTTTCAGTTTCT 1107  
Db GGGGTGCGGAGACGCTGACGCGCGCGGTGCTCGGACGCTCGGTGCTGTGGTTCGAA 1671  
1108 TTTTCCGGTGGAGCAAGCCCTTTTGGCCCACTACCCCGCGCAGCGCAATGTGTG 1167  
Db CCGGACTGGAGGGGAGCGGGCTGCGCGTGGCGCGGAGCGCTGGCGCGGACTGCGCTG 1731

1168 GTGTACAGCGGAAACCCACGACGACACACACCTTGGATGGTTTCGACCGCGC 1222  
Db 1732 GCACCGCCACACACACGACCTGCGCGGACCGCGCGCTACCGGGCGC 1786  
RESULT 13  
US-10-156-761-1  
Sequence 1, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 1  
LENGTH: 9025608  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (4187715)  
OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 5.6%; Score 84.2; DB 6; Length 9025608;  
Best Local Similarity 46.6%; Pred. No. 1.7e-12;  
Matches 389; Conservative 0; Mismatches 428; Indels 18; Gaps 3;  
391 GCCTTTATCGAGCGCGAGCGCTTCTGGCTGGAAGACTATGCGCTCTTTATGGCCCTCAAG 450  
Db GACTTCTCGCGAGAGGGCGAGCGCTGGAGGACCAACGCGAGCTGTGTGGCGCT---G 6537138  
451 ACCCGGTTTGACGCAAGCCCTGGAAGACTGGAGCCCGAGCTGCGGACCGTGAACCG 510  
Db GCCGAGGTGTACGGGGCGGATGGGACGCTGGCGCGCGCGCTGCGGACCGCGCTCG 6537195  
511 GCTGCCCTGCGGAGCGCGCTGAGGAGCTGCGCGAGGAGTGGCCCTTTACGAGTGAATT 570  
Db GCCGAGACCGCGCGCGCGCGGAGTGTGATGACCGCGCTGCGACTTCCAGCGCTGGCTC 6537314  
571 CAGTGGCTTTTATCTGGAATGGGGCGCAGACCAAGCGCTATGCGGAATCCAAGGGGATT 630  
Db GCCTGGCTGACCGACGAGCAGCTGCGCGCGCTCAGCGGGCGCGCGCTGACGCGGGCATG 6537374  
631 CAGATTATCGGCGATATGCCCATCTTTGTGCGCTTCCATTCCTCAGATGTCTGGGCCAAC 690  
Db GCGGTCCGAGCTCGGTGACGACCTTCGCGGTTCGCGGTCCATCCCGTGGCGCGGACGCTCG 6537434  
691 CCGCAGTACTTCTACCTCGAGGCGGATGGCAACCCACGCTGGTGGGGGGTTTCGCGG 750  
Db GCGGACGAGGAGTACTTTCGCGCGGGCA-----TGTCGGTGGGGCGCGCGCC 6537482  
751 GACTACTTCTCGAAACCGCGCAGCTCTGGGGCAATTCGGCTCTATTCGCTGGGATGTGATG 810  
Db GAGCGCTTCAAGCGCTCGGCGAGACTGGGGCTGCGCGCTGCGGACCGGACCGCTC 6537483  
811 GAAAGGCAAACTTTGCTGTGATTTGCCCGCATAGGCGAGTCTCTCAAGCAGTGCAC 870  
Db GCCGAGTCCGGCTACGCGCGCTACCGCGCTGCTGCTCGCGCGCTTCTCCGATACGCGGGC 6537543  
871 CTGGTGGCATCGACCACTTCCGCGGGTGTGAAGCCTACTGGGAGGTTCGTTTGGCGG 930

Db 6537603 GCCCTGCCATCGACCGAGTATGGGCTGTTCCGGCTGTGGTCCCGCAGGGGGG 6537662.  
Qy 931 CCATATGTGTGAGGGGGCGTGGGTCAAGCCGCCAGGGGGAAGCT---GTTTGTCTGG 987  
Db 6537663 CCGCCACGGAGGGAACGTACGTCCGGGTACACGCCGAGGGGATGCTCGCATCTCGTC 6537722  
Qy 988 GTGGGGGCCCAACTGAGCGATGCGCCCATATTCCGAAGACCTGGGGGTGATCAACCCC 1047  
Db 6537723 CTGAGGAGCATCGCCACCGGGGCACTGGTGTATCGGAGGACCTCGGCACGGTTCGAACCC 6537782  
Qy 1048 GAGTGTGAGGCTTTTGGCGGATGGTTCGGGTTCGCCGCGCATGAAGATTTTCAGCTTTGCT 1107  
Db 6537783 GGSGTGGCGAGACGCTGCACCGCGGGCGTGTCTCGCAGTGGTGTCTGAA 6537842  
Qy 1108 TTTTCCGGTGAAGACAAGCCCTTTTGGCCCAACATACCCCGCGCAGGCAATGTGGTG 1167  
Db 6537843 CGGAGCTGGAGGGGAGCGGCTCGCGTCCGCCCGAGCGCTGGCGCGGAGTGCCTG 6537902  
Qy 1168 GTGTACAGCGGAACCCAGCAACACACACCTGGATGGTTCGACACCGGC 1222  
Db 6537903 GCCACCGCACCCACCGACCTGCGCGGACCGCGCGCTCACCGGGCGCC 6537957

## RESULT 14

US-10-424-599-28135  
; Sequence 28135, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 28135  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(563)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_125407C.1  
US-10-424-599-28135

Query Match 5.1%; Score 76.8; DB 7; Length 563;  
Best Local Similarity 53.8%; Pred. No. 4.3e-11;  
Matches 156; Conservative 0; Mismatches 134; Indels 0; Gaps 0;  
Qy 571 CAGTGGCTTTTATCTGGAATGGGGCGAGCAAGCCCTATGCCGAATCCAGGGGATT 630  
Db 252 CAGTTCCTGTTTCAAGCGAGTGGCGAGAAAGTTCCACAGCTATGCACAGAGTAAGGGAATC 311  
Qy 631 CAGATTATCGGGATATGCCCATCTTTGTGGCTTCGATCTTCAGATGTTCTGGGCCAAC 690  
Db 312 AGCAATAAGGAGACATGCCAATATATGTTGGTTATCATAGCGCAGATGTTGGGCAAT 371  
Qy 691 CCGCAGTACTTCTACCTCGAGCGCGATGGCAACCCCGGCGGTGGGGGGGTTCCCGCG 750  
Db 372 AAGAAACAGTTTCTGACAGGAGGGCTTTCTCTTTTAGTTAGTGGTGTCTCTCT 431  
Qy 751 GACTACTTCTCCGAAACCGGCGAGCTCTGGGGCAATCGGCTCTATCGCTGGGATGATG 810  
Db 432 GACGCAATTCAGTGAACCTGGTCAGCTTTGGGGCAGCCCTGTATGATGGAAAGCAATG 491  
Qy 811 GAAAGGACAACTTTGGCTGGTGTGATTTGCCCGCATAGGCAATCGCTCAA 860

Db 492 GAGAAAGATGGATACCTCATGGGGTACGTNGCATACGACGNGCACAAAA 541

## RESULT 15

US-10-425-114-29452  
; Sequence 29452, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 29452  
; LENGTH: 1945  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-GMFLMINSOY045C05\_FLI  
US-10-425-114-29452

Query Match 4.4%; Score 66.8; DB 7; Length 1945;  
Best Local Similarity 57.0%; Pred. No. 3.2e-08;  
Matches 122; Conservative 0; Mismatches 92; Indels 0; Gaps 0;  
Qy 571 CAGTGGCTTTTATCTGGAATGGGGCGAGCAAGCCCTATGCCGAATCCAGGGGATT 630  
Db 326 CAGTCTTGTGTTTCAAGCGAGTGGCGAGAAAGTTTCATAGCTATGCACAGTAAGGGAATC 385  
Qy 631 CAGATTATCGGGATATGCCCATCTTTGTGGCTTCGATTCCTCAGATGTTCTGGGCCAAC 690  
Db 386 AGCATATGGAGACATGCCAATATATGTTGTTATCATAGCGCAGATGTTGGGCAAT 445  
Qy 691 CCGCAGTACTTCTACCTCGAGGGCGGATGGCAACCCCGGCGGTGGGGGGGTTCCCGG 750  
Db 446 AAGAAACAGTTTCTACTGAAACAGAGGGGCTTTCTCTTTTAGTTAGTGTGTTCTCTCT 505  
Qy 751 GACTACTTCTCCGAAACCGGCGAGCTCTGGGGCA 784  
Db 506 GACGCATTTCAGTGAACCCCGGCGAGCTCTGGGGCA 539

Search completed: January 14, 2006, 17:18:12  
Job time : 1164 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2006, 14:44:31 ; Search time 398 Seconds  
(without alignments)  
3056.409 Million cell updates/sec

Title: US-10-618-976-1

Perfect score: 1503

Sequence: 1 atgcaactccaacgcgcttt.....ccgaggccagccagcgct 1503

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6038814 seqs, 404674181 residues

Total number of hits satisfying chosen parameters: 12077628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New:

- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq\*
- 4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq\*
- 5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq\*
- 6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq\*
- 7: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq\*
- 8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq2\*
- 9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq3\*
- 10: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match %	Score	Length	DB ID	Description
1	43	2.9	88421	7	US-11-205-109-1
2	40.2	2.7	1647	7	US-11-137-465-14
3	39.6	2.6	761	6	US-10-497-135-18
C 4	39	2.6	88421	7	US-11-205-109-1
5	38.6	2.6	3618	7	US-11-136-527-294
6	38.2	2.5	1400	7	US-11-136-527-4370
7	38.2	2.5	6021	7	US-11-136-527-274
C 8	38	2.5	2749	7	US-11-000-688-100
9	37.6	2.5	2149	7	US-11-078-189-6
10	37.2	2.5	10968	7	US-11-075-185-35
11	37.2	2.5	78869	7	US-11-075-185-1
12	37.2	2.5	116856	7	US-11-143-980-1
13	37.2	2.5	168516	7	US-11-121-086-3
14	36.8	2.4	10524	7	US-11-075-185-37
15	36.6	2.4	1987	7	US-11-128-061-384
16	36.4	2.4	767	6	US-10-497-135-17
C 17	36.4	2.4	7006	6	US-10-821-234-218
18	35.6	2.4	864	7	US-11-179-411-26
19	35.6	2.4	864	7	US-11-175-766-26
C 20	35.6	2.4	1974	6	US-10-821-234-34
21	35.6	2.4	11070	7	US-11-075-185-34
C 22	35.6	2.4	37507	6	US-10-522-037-2
23	35.4	2.4	2023	6	US-10-995-561-55

24	35.4	2.4	2391	7	US-11-136-527-1926
C 25	35.4	2.4	11684	6	US-10-995-561-498
26	35.4	2.4	12591	6	US-10-995-561-13415
27	35.4	2.4	14023	6	US-10-995-561-13221
C 28	35.4	2.4	199130	6	US-10-995-561-13223
29	35.2	2.3	930	6	US-10-524-647-127
30	35.2	2.3	2570	7	US-11-136-527-3380
31	35	2.3	1290	7	US-11-136-527-2457
32	35	2.3	1290	7	US-11-136-527-6553
C 33	35	2.3	116856	7	US-11-143-980-1
34	34.8	2.3	2025	7	US-11-167-048-2
35	34.8	2.3	2736	6	US-10-858-730-38
36	34.8	2.3	5679	7	US-11-075-185-36
C 37	34.8	2.3	15515	7	US-11-112-908-42
C 38	34.8	2.3	159660	7	US-11-112-908-43
C 39	34.8	2.3	177623	7	US-11-112-908-41
C 40	34.6	2.3	609	6	US-10-467-657-8011
C 41	34.6	2.3	690	6	US-10-858-730-156
42	34.6	2.3	6941	6	US-10-432-483-49
43	34.6	2.3	179597	7	US-11-121-086-91
44	34.4	2.3	489	7	US-11-120-308-5
45	34.4	2.3	543	7	US-11-120-308-11

#### ALIGNMENTS

RESULT 1  
US-11-205-109-1  
; Sequence 1, Application US/11205109  
; Publication No. US20050287641A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Zazopoulos, Emmanuel  
; APPLICANT: Staffa, Alfredo  
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS  
; FILE REFERENCE: 3002-2US  
; CURRENT APPLICATION NUMBER: US/11/205,109  
; PRIOR FILING DATE: 2005-08-17  
; PRIOR FILING DATE: 2001-10-15  
; PRIOR FILING DATE: 2001-10-15  
; PRIOR FILING DATE: 2000-10-13  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 88421  
; TYPE: DNA  
; ORGANISM: Actinoplanes sp.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2077)..(3078)  
; OTHER INFORMATION: ORF 1; positive strandedness  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (3118)..(4032)  
; OTHER INFORMATION: ORF 2; positive strandedness  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4038)..(5048)  
; OTHER INFORMATION: ORF 3; positive strandedness  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (6665)..(5814)  
; OTHER INFORMATION: ORF 4; negative strandedness  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (7703)..(6693)  
; OTHER INFORMATION: ORF 5; negative strandedness  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (9464)..(8130)  
; OTHER INFORMATION: ORF 6; negative strandedness





```

; PRIOR APPLICATION NUMBER: US/10/239,663
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/09226
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/192,158
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/192,668
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/200,166
; PRIOR FILING DATE: 2000-04-27
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 1647
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-137-465-14

Query Match      2.7%; Score 40.2; DB 7; length 1647;
Best Local Similarity 48.5%; Pred. No. 2.5;
Matches 111; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 1254 CTTGCCCGCTATGCGATCCGTTGTTTGTTCGGAATACAGGTGCGGGCGCTTTGATGGA 1313
Db      |||||
Qy 336 CGTGTGCGCTGCCCGACCGCTGTGAGAGGAGCCAGCTTCACCTCGCGCTCGGACGG 395
Db      |||||
Qy 1314 CTTGCCCTTCAAGCCCGCCAGCTGGCTATTGTCCTTTGAGGACGCTGCTGGGGCT 1373
Db      |||||
Qy 396 CCTCACCTACTACACCGCTCTATATGAGCGCGAGCGCTGCTGCGGGCGCTGCACT 455
Db      |||||
Qy 1374 GGGGCGCGAGCGCGCATGAATCTCCCGGACGCTGGGGGACAACTGGGCGTGGCGCTA 1433
Db      |||||
Qy 456 CCAATGTCGCTGCAAGACGCTCTAGTGGCGGCCAGACCGCGGGCGCGCGGA 515
Db      |||||
Qy 1434 CGCCGAAGGCGACTCGAGCGCGCTGTCGCGCGGGAATGCGGGCGCTG 1482
Db      |||||
Qy 516 GACCACTGCGCGCGCCACACCTGCGGGCGCGCGCTGCTCTGCTGCTG 564
Db      |||||

RESULT 3
US-10-497-135-18
; Sequence 18, Application US/10497135
; Publication No. US2005027132A1
; GENERAL INFORMATION:
; APPLICANT: Biotica Technology Ltd
; APPLICANT: Gregory, Matthew A
; APPLICANT: Gaisner, Sabine
; APPLICANT: Petkovic, Hrvoje
; APPLICANT: Moss, Steven
; TITLE OF INVENTION: Production of Polyketides and Other Natural Products
; FILE REFERENCE: 4408-P03444US00
; CURRENT APPLICATION NUMBER: US/10/497,135
; CURRENT FILING DATE: 2004-05-26
; PRIOR APPLICATION NUMBER: PCT/GB03/003230
; PRIOR FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: GB0216509.0
; PRIOR FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: GB0224922.5
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 761
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-497-135-18

Query Match      2.6%; Score 39.6; DB 6; Length 761;
Best Local Similarity 47.7%; Pred. No. 3.1;
Matches 148; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

Qy 1117 GAGGACAAAGCGCTTTTGTGCCCCACAACTACCCCGCGACGCAATGTGGTGTACAGC 1176
Db      |||||

```



```
FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3524)..(3525)
; OTHER INFORMATION: n is a, c, g, or t
US-11-136-527-294

Query Match      2.6%; Score 38.6; DB 7; Length 3618;
Best Local Similarity 50.3%; Pred. No. 6.4; Indels 0; Gaps 0;
Matches 95; Conservative 0; Mismatches 94;

Qy 366 GSCCCAGGATAAGACCCGACTTGAAGCCTTTATCGAGCCCGAGCGCTTCTGGCTGGAGA 425
Db 366 GGAAGAAGATTGAGAAGAGCTGGAGCGCTGTGCCAGGACGTCTGAGCCTGCTGGACAA 425
Qy 426 CTATGCGCTTTTATGGCCCTCAAGACCCGCTTTTACGGGCAAGCCCTGGACGAGTGGAG 485
Db 426 CTACCTGATCAAGAAGCTGCAGCGAGACCCAGTACGAGAGCAAGGTGTTCTACCTGAAGAT 485
Qy 486 CCCCAGAGCTGGCGACCGTGAACCGCTGCCCTGGCCAGGCGCCGTGAGGAGCTGGCCGA 545
Db 486 GAAAGGGGACTATTACCCCTACTCTGGCTGAAGTGGCCACTGGGGAGNAAGGGCGACCGT 545
Qy 546 GGAGGTGGC 554
Db 546 GGTGAGTC 554

RESULT 6
US-11-136-527-4370
; Sequence 4370, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4370
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-4370

Query Match      2.5%; Score 38.2; DB 7; Length 1400;
Best Local Similarity 39.4%; Pred. No. 7; Indels 0; Gaps 0;
Matches 98; Conservative 27; Mismatches 124;

Qy 400 GAGCCGAGCGCTTCTGGCTGGAAGACTATGCGCTCTTTATGGCCCTCAAGACCCGGTTT 459
Db 11 GAGAACAGAAACCTCCAGAGAGATCTCAGACCTGACTGACAGCTGGGHHMBASBKGGK 70
Qy 460 GACGGCAAGCCCTGGAACGAGTGGAGCCCGAGCTGCGGACCGGTGAACCGGCTGCCCTG 519
Db 71 WARABRTSCAGAGCTGAGAGATCCGMAARCARCTGGAGGYKGADMMKCTSSWMTG 130
Qy 520 GCCAGGGCCCTGAGGAGCTGCCGAGAGAGTGGCCCTTTACGAGTGAATTCAGTGGCTT 579
Db 131 CMGTGAGCCCTGGAGGAGGCTTCGAGGATGAGGAGGCGCAAGATCCTCCGA 190
Qy 580 TTTTATCTGGAATGGGCGCAGACCAAGCCCTATGCCAATCCAAAGGGGATTCAGATTATC 639
Db 191 GCYAGCTGGAGTTCAMCCAGATYAAGGSAGAGATYGAAAGAGAGCTGGCAGAGAAGAC 250
Qy 640 GGCGATATG 648
Db 251 GAGGAGATK 259

FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2749)
; OTHER INFORMATION: colony stimulating factor 1
; OTHER INFORMATION: (macrophage) (CSF1) gene.
```

```
RESULT 7
US-11-136-527-274
; Sequence 274, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 274
; LENGTH: 6021
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-274

Query Match      2.5%; Score 38.2; DB 7; Length 6021;
Best Local Similarity 39.4%; Pred. No. 8.4; Indels 0; Gaps 0;
Matches 98; Conservative 27; Mismatches 124;

Qy 400 GAGCCGAGCGCTTCTGGCTGGAAGACTATGCGCTCTTTATGGCCCTCAAGACCCGGTTT 459
Db 4632 GAGAACAGAAACCTCCAGAGAGATCTCAGACCTGACTGACAGCTGGGHHMBASBKGGK 4691
Qy 460 GACGGCAAGCCCTGGAACGAGTGGAGCCCGAGCTGCGGACCGGTGAACCGGCTGCCCTG 519
Db 4692 WARABRTSCAGAGCTGAGAGATCCGMAARCARCTGGAGGYKGADMMKCTSSWMTG 4751
Qy 520 GCCAGGGCCCTGAGGAGCTGGCGGAGGAGTGGCCCTTTACGAGTGAATTCAGTGGCTT 579
Db 4752 CMGTGAGCCCTGGAGGAGGCTTCGAGGATGAGGAGGCGCAAGATCCTCCGA 4811
Qy 580 TTTTATCTGGAATGGGCGCAGACCAAGCCCTATGCCAATCCAAAGGGGATTCAGATTATC 639
Db 4812 GCYAGCTGGAGTTCAMCCAGATYAAGGSAGAGATYGAAAGAGAGCTGGCAGAGAAGAC 4871
Qy 640 GGCGATATG 648
Db 4872 GAGGAGATK 4880

FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2749)
; OTHER INFORMATION: colony stimulating factor 1
; OTHER INFORMATION: (macrophage) (CSF1) gene.
```

```
RESULT 8
US-11-000-688-100/c
; Sequence 100, Application US/11000688
; Publication No. US20050287544A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, Francois
; APPLICANT: HOULGATTE, Remi
; APPLICANT: BIRNBAUM, Daniel
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF COLON CANCER WITH DNA ARRAYS
; FILE REFERENCE: 1423-R-03
; CURRENT APPLICATION NUMBER: US/11/000,688
; CURRENT FILING DATE: 2004-12-01
; PRIOR APPLICATION NUMBER: US 60/525,987
; PRIOR FILING DATE: 2003-12-01
; NUMBER OF SEQ ID NOS: 1596
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 100
; LENGTH: 2749
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial sequences:primer
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2749)
; OTHER INFORMATION: colony stimulating factor 1
; OTHER INFORMATION: (macrophage) (CSF1) gene.
```

```
US-11-000-688-100
Query Match      2.5%; Score 38; DB 7; Length 2749;
Best Local Similarity 51.9%; Pred. No. 8.5;
Matches 109; Conservative 0; Mismatches 100; Indels 1; Gaps 1;

QY 1286 AATACAGAGTGGGGCGCTTTGATCGAGTGGCTTCAAAAGCCCGGCGCAAGCTGGCTA 1345
DB      |||||
QY 1346 TTGTGCTTTGAGAGAGCTGCTGGGCTGGGCGCCCGAGGC-CGGCATGAATTCCTCCCGGA 1404
DB      |||||
QY 1405 CGGCTGGGGGACAACTGGGCGTGGCGCTAGCCGAAAGGCGACCTCGAGCCCGGTCTGGCC 1464
DB      |||||
QY 1465 GCGGAGTGGGCGCTTGGCGGAGGCGCCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 1525
DB      |||||

RESULT 9
US-11-078-189-6
; Sequence 6, Application US/11078189
; Publication No. US2005027167A1
; GENERAL INFORMATION:
; APPLICANT: Bachinger, Hans Peter
; APPLICANT: Vranka, Janice
; TITLE OF INVENTION: PROLYL 3-HYDROXYLASES
; FILE REFERENCE: 08062-020001
; CURRENT APPLICATION NUMBER: US/11/078,189
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: US 60/552,409
; PRIOR FILING DATE: 2004-03-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 2149
; TYPE: DNA
; ORGANISM: Mus musculus
US-11-078-189-6

Query Match      2.5%; Score 37.6; DB 7; Length 2149;
Best Local Similarity 47.8%; Pred. No. 10;
Matches 109; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 1225 GAGCGGAGCGGGCTTTCATGCGGGGCTTACCTGGCGCGCTATGGCATCGCTGTTTGTG 1284
DB      |||||
QY 1285 GAATACAGAGTGGGCGCTTTGATCGAGTGGCTTCAAAAGCCCGGCGCAAGCTGGCT 1344
DB      |||||
QY 310 GGGCGCGGGGTCTGGGAACCGCTGCTCTGGCGGCGACGCTCGCGCGCGCGCGCTCG 369
QY 1345 ATTGTGCTTTGAGAGAGCTGCTGGGCTGGGCGTGGGCGCGCGCGCGCGCTCGCGGAGC 1404
DB      |||||
QY 1405 CGGCTGGGGGACAACTGGGCGTGGCGTACCGCGAAGGCGACCTCGAG 1452
DB      |||||
QY 430 CGCGTGGGGATGCTTCCGCGCGGCGGAGCTTACAACTACCTGCAG 477
DB      |||||

RESULT 10
US-11-075-185-35
; Sequence 35, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH

Query Match      2.5%; Score 37.2; DB 7; Length 7869;
Best Local Similarity 45.7%; Pred. No. 20;
Matches 129; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 1183 CAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1242
DB      |||||
QY 22357 CGCGCCCTGTAGCCACCCAGCCCGCTTCCGCGCGCGCGCTCGAGCCGCGCTCGCGCCG 22416
DB      |||||

US-11-075-185-1
; Sequence 1, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 7869
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-11-075-185-1

Query Match      2.5%; Score 37.2; DB 7; Length 10968;
Best Local Similarity 45.7%; Pred. No. 16;
Matches 129; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 1183 CAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1242
DB      |||||
QY 7090 CGCGCCCTGTAGCCACCCAGCCCGCTTCCGCGCGCGCGCTCGAGCCGCGCTCGCGCCG 7149
DB      |||||
QY 1243 ATGCGGGCTACCTGCGCGCTATGGCATCCGTTGTTGTCGGAATACAGAGTTCGCGGC 1302
DB      |||||
QY 7150 CTCGACCGGACCTCGACCGGCGCTCATGAGCGCTCTTCGCGGAGCCGCGCTCGGAG 7209
DB      |||||
QY 1303 GCTTTGATCGAGCTGGCTTCAAAAGCCCGGCGGCAAGCTGGCTATGTCCTTGGCAGGAC 1362
DB      |||||
QY 7210 GCGCGGGCTGCTCGACGAGCAGCAGCTGGGCTCGAGCCGCGCTCTTCGCTTCGAGGTG 7269
DB      |||||
QY 1363 GTGCTGGGCTGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1422
DB      |||||
QY 7270 GCGCTTACCGGCGGAGTGGGAGTGGGCGCTGGGCGCTGGGCGGCGGCGGCGGCGGCGG 7329
DB      |||||
QY 1423 GCGTGGCGCTACGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1464
DB      |||||
QY 7330 ATCGGGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7371
DB      |||||

RESULT 11
US-11-075-185-1
; Sequence 1, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 7869
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-11-075-185-1

Query Match      2.5%; Score 37.2; DB 7; Length 7869;
Best Local Similarity 45.7%; Pred. No. 20;
Matches 129; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 1183 CAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1242
DB      |||||
QY 22357 CGCGCCCTGTAGCCACCCAGCCCGCTTCCGCGCGCGCGCTCGAGCCGCGCTCGCGCCG 22416
DB      |||||
```

	Query Match	2.5%	Score 37.2;	DB 7;	Length 168516;
	Best Local Similarity	46.4%;	Pred. No. 22;		
	Matches 156;	Conservative	0;	Mismatches 178;	Indels 2; Gaps 1;
Qy	1110	TTCCGCTGAGGCAACGCGCTTTTGGCCCCACAACTACCCCGCAGCGCAATGFGTGGT	1169		
Db	135213	TGCGCTGACCTTCATGTCATGGCACAGATCTCCATGGCGCCACAGCCCTCGCGATGG	135272		

DB	135273	ACACAGCCCCGCTGAGCACGGCCAGAGCCCGTGGGATGGAGCACTCGAGCGCGAGGC	135332
QY	1230	CGAGCGGGCGCTTCATGCGGGCGCTACTGGCGCGCTATGGCATCGGTTGTTTGTCTCGAATA	1289
DB	135333	AGAGCCAGTTCGACAGAGAGGGCGGGGTCGGCAGAGCAGAGCCCTGGCGGGCCACTGAGGCCA	135392
QY	1290	CGAGGTTCGGCGGGCGCTTTGATTCAGCTGGCGCTTCAAAAGCCCGGCCAAGCTGGCTATTGT	1349
DB	135393	GTTCGCAGCTGAGCCTTGGCGGGACGGGCGAGAAAGGGCTGCGAGGGGCTGGCTTCCT	135452
QY	1350	GCCTTTGCAGGACGTCCTGGGCTGGCGCCCGG--GGCCCGCATGACTTCCCGGACGG	1407
DB	135453	GGAGTGCTGGGAGCGCCCGAGGCGATGAACGAGCGACCCACATTCAGCACCTCTGGGTG	135512
QY	1408	CTGGGGGACAACCTGGCGCTGGCGCTACGCGAAGGC	1443
DB	135513	CTGGGGAGCACAAGTGGCAGTGCCAGGCGTTGGCG	135548

```

/ CURRENT APPLICATION NUMBER: US/11/075,185
/
/ CURRENT FILING DATE: 2005-03-07
/
/ PRIOR APPLICATION NUMBER: US 60/551,103
/
/ PRIOR FILING DATE: 2004-03-08
/
/ PRIOR APPLICATION NUMBER: US 60/568,290
/
/ PRIOR FILING DATE: 2004-05-04
/
/ NUMBER OF SEQ ID NOS: 61
/
/ SOFTWARE: PatentIn version 3.3
/
/ SEQ ID NO 37
/
/ LENGTH: 10524
/
/ TYPE: DNA
/
/ ORGANISM: Sorangium cellulosum
/
/ US-11-075-185-37

```

[illegible]

Tue Jan 17 09:28:40 2006

QY 1450 GAGCCGGCTCTGCGCGGACTGCGGGCCCT 1481  
| | | | | | | | | | | | | | | | | |  
Db 5059 GCGAGCGCAAGGCTCGGGCTGCGGCGCT 5090

## RESULT 15

US-11-128-061-384  
; Sequence 384, Application US/11128061  
; Publication No. US2006003958A1  
; GENERAL INFORMATION:  
; APPLICANT: Melville, Mark W.  
; APPLICANT: Charlebois, Timothy S.  
; APPLICANT: Mounts, William M.  
; APPLICANT: Hann, Louane E.  
; APPLICANT: Sinacore, Martin S.  
; APPLICANT: Leonard, Mark W.  
; APPLICANT: Brown, Eugene L.  
; APPLICANT: Miller, Christopher P.  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS  
; TITLE OF INVENTION: TO MONITOR GENE EXPRESSION  
; FILE REFERENCE: 01997.027701  
; CURRENT APPLICATION NUMBER: US/11/128,061.  
; CURRENT FILING DATE: 2005-05-11  
; PRIOR APPLICATION NUMBER: US 60/570,425  
; PRIOR FILING DATE: 2004-05-11  
; NUMBER OF SEQ ID NOS: 7285  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 384  
; LENGTH: 1987  
; TYPE: DNA  
; ORGANISM: Mesocricetus auratus  
US-11-128-061-384

Query Match 2.4%; Score 36.6; DB 7; Length 1987;  
Best Local Similarity 50.3%; Pred. No. 17;  
Matches 90; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 927 CCGGCCCAATGCTGTGAGGGGGCTGGTCAAGCCCCCAGGGGAGAGCTGTTTGTCTGC 986  
| | | | | | | | | | | | | | | | | |  
Db 1342 CCTGATGATTGGTCTCATCTGGCGCTCCATGGAACACCCAGGGAAGCTCCTGTTGTCTCC 1401  
QY 987 GGTGCGGGCCCACTGAGCGATGCGCCCATCATTCGCCAGACCTGGGGGTGATCACCCC 1046  
| | | | | | | | | | | | | | | | | |  
Db 1402 TAAATTTGCTCCTGGACAGGAATCAGGGCAAGTGTGTGGAGGGCATGGTGAGATCTTTGA 1461  
QY 1047 CGAGTGGAGGCTTTGCGCGATGCTTCGGGTCCCGGATGAAGATTTTGCAGTTTG 1105  
| | | | | | | | | | | | | | | | | |  
Db 1462 CATGTTGCTGCTACATCAGCTCGTTCCGATGATGACTGCGAGGAGAGGTTTG 1520

Search completed: January 14, 2006, 17:24:45  
Job time : 400 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2006, 06:57:40 ; Search time 880 Seconds  
(without alignments)

11382.992 Million cell updates/sec

Title: US-10-618-976-1

Perfect score: 1503

Sequence: 1 atgcaactccaagcgcttt.....ccgaggccagccagcgct 1503

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq\_21.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002as.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004as.\*
- 13: Geneseqn2004bs.\*
- 14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1503	100.0	1503	5	Ad04867 Thermus r
2	638.4	42.5	1503	2	AAV7539 Thermus f
3	638.4	42.5	2085	12	AdQ15010 DNA encod
4	636.8	42.4	1503	13	AdC23306 DNA of Th
5	378	25.1	1539	13	AdU07728 DNA seque
6	329.2	21.9	1518	12	AdQ15025 DNA encod
7	289	19.2	1518	13	AdK45052 Streptoco
8	287.4	19.1	1515	10	Abx07913 S. pneumo
9	287.4	19.1	5910	2	AAV52298 Streptoco
10	287.4	19.1	110000	10	ABSS56454 19
11	265.6	17.7	1377	13	AdR91701 Novel S.
12	265.6	17.7	1377	13	AEAS5571 Streptoco
13	181.4	12.1	1731	12	AdQ15021 DNA encod
14	173	11.5	1491	6	ABN69262 Streptoco
15	169.4	11.3	1494	6	ABN69261 Streptoco
16	169.4	11.3	1497	13	AdV84037 Streptoco
17	169.4	11.3	16857	13	AdV87699 Streptoco
18	169.4	11.3	16857	13	AdV78952 Streptoco
19	169.4	11.3	110000	6	ABN71527_14

c	20	169.4	11.3	110000	13	ADV81204_15
	21	164	10.9	1581	12	ADQ15027
	22	160.6	10.7	1458	10	ADC23308
	23	159	10.6	1676	13	ADK60464 Plant ful
	24	113.2	7.5	1485	14	ACL66091 M. xanthu
c	25	113.2	7.5	15377	14	ACL64653 M. xanthu
	26	91.4	6.1	110000	2	AAAX20248_07
	27	66.8	4.4	1945	13	ADX54712 Plant ful
	28	64.6	4.3	259	7	ADS65998 Corn seed
	29	64.2	4.3	1557	3	AAZ45687 DNA encod
	30	58.6	3.9	2097	6	ABA99452 Actinopia
	31	58.6	3.9	38064	6	ABA99469 Actinopia
	32	56.8	3.8	2217	6	ABZ12444 Arabidops
	33	56.2	3.7	459	4	AA07075 DNA encod
	34	55.6	3.7	978	10	ABZ66744 Orthosomy
	35	55.6	3.7	45055	10	ABZ66808 Orthosomy
c	36	54	3.6	2000	8	ADA71938 Rice gene
	37	50	3.3	36401	14	ADY26621 Streptomy
	38	49.4	3.3	1959	13	ADT17699 Plant CDN
	39	49.4	3.3	2000	8	ADA71938 Rice gene
	40	48.8	3.2	745	10	ADG89856 Corn 4-al
	41	48.6	3.2	745	2	AAV70279 Corn 4-al
	42	48.2	3.2	921	11	ACL33682 Rice abio
	43	48.2	3.2	1146	10	ABT23235 Seed deve
	44	47.6	3.2	2658	8	ACA37790 Prokaryot
c	45	47.4	3.2	1836	5	AAH67448 C glutami

#### ALIGNMENTS

##### RESULT 1

Ad04867 AAD04867 standard; DNA; 1503 BP.

XX AAD04867;

DT 06-AUG-2003 (revised)

DT 17-JUL-2001 (first entry)

XX Thermus rubens glucanotransferase gene.

XX Glucanotransferase; starch; Japanese dessert; noodles; wheat product;  
XX sea food; processed food; baby; animal; drink; gyoza skin; shumai skin;  
XX nutrient supplemental food; snack; detergent composition; ds.

XX Thermus sp.

XX Key Location/Qualifiers

FT CDS 1..1503

FT /\*tag= a

FT /product= "Thermus rubens glucanotransferase"

FT /notes= "Does not include stop codon"

FT /partial

XX WO200129195-A1.

XX 26-APR-2001.

XX 16-OCT-2000; 2000WO-DK000560.

XX 20-OCT-1999; 99DK-00001501.

XX 15-NOV-1999; 99DK-00001641.

XX (NOVO ) NOVOZYMES AS.

XX Fukuyama S;

XX WPI; 2001-290912/30.

XX P-PSDB; AAE01153.

XX Novel polypeptides having glucanotransferase activity for producing food  
PT and in detergent or cleaning compositions for removing starch stains,

PT	especially amylose stains.	
XX	Claim 16; Page 49-52; 57pp; English.	
PS	The present DNA sequence is <i>Thermus rubens</i> ATCC 31556 glucanotransferase	
XX	gene. Glucanotransferase catalyzes the cyclisation of starch by acting on	
CC	starch in food to produce a cyclic glucan. Glucanotransferase is useful	
CC	for producing food such as Japanese desserts, snacks, wheat products,	
CC	noodles, gyoza skins, shumai skins, processed sea foods, frozen or	
CC	refrigerated processed foods, weaning foods, baby foods, pet foods,	
CC	animal feeds, drinks, sports foods and nutrient supplemental foods. A	
CC	cleaning or detergent composition comprising glucanotransferase is useful	
CC	for removal of starch stains, in particular for removal of amylose	
CC	stains, from a hard surface or from laundry, by contacting the amylose	
CC	stain-containing hard surface or the amylose stain-containing laundry	
CC	with glucanotransferase or its composition. Nucleic acid constructs,	
CC	vectors and host cells comprising glucanotransferase nucleic acid	
CC	sequence are also provided. (Updated on 06-AUG-2003 to correct OS field.)	
XX		
SQ	Sequence 1503 BP; 248 A; 452 C; 497 G; 306 T; 0 U; 0 Other;	
	Query Match 100.0%; Score 1503; DB 5; Length 1503;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 1503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 ATGCACTCCAAACCGCTTTTGGAAATTTGCTCACCCACACAGTTTTCGGGTGCGTGG 60	
Db	1 ATGCACTCCAAACCGCTTTTGGAAATTTGCTCACCCACACAGTTTTCGGGTGCGTGG 60	
Qy	61 GGGATTGGGGCTTGGGCGCGGAGCGGCTTTTGGACTGGCTGGCGGATGCGGGA 120	
Db	61 GGGATTGGGGCTTGGGCGCGGAGCGGCTTTTGGACTGGCTGGCGGATGCGGGA 120	
Qy	121 GCCCGCTGGTGGAGGCTTACCGCTGGGCGCTACACAGTTACGGCGACTCGCGTACCAG 180	
Db	121 GCCCGCTGGTGGAGGCTTACCGCTGGGCGCTACACAGTTACGGCGACTCGCGTACCAG 180	
Qy	181 TCCTTCTCGGCTTTTCCGGTAAACCCGTAATTTGGTTGACCCCGAGATGCTGATGAAAAA 240	
Db	181 TCCTTCTCGGCTTTTCCGGTAAACCCGTAATTTGGTTGACCCCGAGATGCTGATGAAAAA 240	
Qy	241 GGCTGGCTGGNACAAAGGAGCGCCCGCGCTATCCGACCGAGCGGTGGATTATGGC 300	
Db	241 GGCTGGCTGGNACAAAGGAGCGCCCGCGCTATCCGACCGAGCGGTGGATTATGGC 300	
Qy	301 TGGCTTTACAGACCCGCTGGCCCTGTGTGGCGGGCTTTTCGGGGGTTTCGGGCAAGG 360	
Db	301 TGGCTTTACAGACCCGCTGGCCCTGTGTGGCGGGCTTTTCGGGGGTTTCGGGCAAGG 360	
Qy	361 GTTTCGGCCCAAGATAAGACCCGACTGGAAGCCCTTTATCGAGGCGGCTTCTGGCTG 420	
Db	361 GTTTCGGCCCAAGATAAGACCCGACTGGAAGCCCTTTATCGAGGCGGCTTCTGGCTG 420	
Qy	421 GAAGACTATGGCTTTTATGGCCCTCAAGACCGGTTTGGACGCGCAAGCCCTGGAACGAG 480	
Db	421 GAAGACTATGGCTTTTATGGCCCTCAAGACCGGTTTGGACGCGCAAGCCCTGGAACGAG 480	
Qy	481 TGGAGCCCGAGCTGCGGACCGTGAAACCGGCTGCCCTTGGCCAGGCGCCGTGAGGAGCTG 540	
Db	481 TGGAGCCCGAGCTGCGGACCGTGAAACCGGCTGCCCTTGGCCAGGCGCCGTGAGGAGCTG 540	
Qy	541 GCCGAGGAGGTGGCCCTTTACAGTGGATTTCAGTGGCTTTTATCTGGAATGGGGCCAG 600	
Db	541 GCCGAGGAGGTGGCCCTTTACAGTGGATTTCAGTGGCTTTTATCTGGAATGGGGCCAG 600	
Qy	601 ACCAAGGCTATGCGGAATCAAGGGGATTCAGATTATCGCGGATATGCCCATCTTTGTG 660	
Db	601 ACCAAGGCTATGCGGAATCAAGGGGATTCAGATTATCGCGGATATGCCCATCTTTGTG 660	
Qy	661 GCCTTCGATTCTCAGATGCTCGGCGCAACCGCAGTACTTCTACCTCGAGGCCGATGGC 720	
Db	661 GCCTTCGATTCTCAGATGCTCGGCGCAACCGCAGTACTTCTACCTCGAGGCCGATGGC 720	

721 RACCCACGGTGGTGGCGGGCGTTCCGGGGGACTACTTCTCCGAAACCGGCCAGCTCTGG 780

721 AACCCACCGTGGTGGCGGGCGTTCCGGGGGACTACTTCTCCGAAACCGGCCAGCTCTGG 780

781 GGCATATCCGCTCTATCGCTGGGATGTGATGGAAGGACAACTTTGCTGCTGGTGCATGGC 840

781 GGCATATCCGCTCTATCGCTGGGATGTGATGGAAGGACAACTTTGCTGCTGGTGCATGGC 840

841 CGCATAGGCACTCGCTCAAGCAGTGCACCTGGTGGCATGACCACTTCGCGGGGTTT 900

841 CGCATAGGCACTCGCTCAAGCAGTGCACCTGGTGGCATGACCACTTCGCGGGGTTT 900

901 GAAGCCTACTCGGAGGTTCCGTTTGGCCCGGCCCAATGCTGTGGAGGGGGCTGGGTCAA 960

901 GAAGCCTACTCGGAGGTTCCGTTTGGCCCGGCCCAATGCTGTGGAGGGGGCTGGGTCAA 960

961 GCCCAGGAGGAAAGCTGTTTGTGCGGTGGGGGCCCAACTGAGCGATGGGCCCATCTT 1020

961 GCCCAGGAGGAAAGCTGTTTGTGCGGTGGGGGCCCAACTGAGCGATGGGCCCATCTT 1020

1021 GCCGAGACCTGGGGTGTATCAACCCCGAGGTGGAGGCTTTGGCGCATGGCTTCGGGTTT 1080

1021 GCCGAGACCTGGGGTGTATCAACCCCGAGGTGGAGGCTTTGGCGCATGGCTTCGGGTTT 1080

1081 CCGGCGATCAAGATTTTGCAGTTTGTCTTTTCCGGTGAGGACAAACGCTTTTTCGCCCTAC 1140

1081 CCGGCGATCAAGATTTTGCAGTTTGTCTTTTCCGGTGAGGACAAACGCTTTTTCGCCCTAC 1140

1141 AACTACCCCGCGCACGGCAATGTGTGTGTACAGGGGAAACCCACGACAAACGACCAAC 1200

1141 AACTACCCCGCGCACGGCAATGTGTGTGTACAGGGGAAACCCACGACAAACGACCAAC 1200

1201 CTGGGATGTCTGGCACCGCGCGGAGCGGCGCTTTCATGCGGGCTTACCTGGCC 1260

1201 CTGGGATGTCTGGCACCGCGCGGAGCGGCGCTTTCATGCGGGCTTACCTGGCC 1260

1261 CGTATGGCATCGCTTGTGTGTGTACAGAGTTCGGGGCGCTTTGATCGAGCTGGCC 1320

1261 CGTATGGCATCGCTTGTGTGTGTACAGAGTTCGGGGCGCTTTGATCGAGCTGGCC 1320

1321 TTCAAAGCCCGCCCAAGCTGCTATTTGCTTTTCAGGACGCTGCTGGGGCTGGGGCCC 1380

1321 TTCAAAGCCCGCCCAAGCTGCTATTTGCTTTTCAGGACGCTGCTGGGGCTGGGGCCC 1380

1381 GAGGCGCGATGACTTCCCGGAGCGCTGGGGGACAACTGGGCGTGGCGCTACGCCGAA 1440

1381 GAGGCGCGATGACTTCCCGGAGCGCTGGGGGACAACTGGGCGTGGCGCTACGCCGAA 1440

1441 GCGGACCTCGAGCCCGTCTGCGCGCGGAGTCTGGGGCGCTGGCGCGCTAGCCGCGC 1500

1441 GCGGACCTCGAGCCCGTCTGCGCGCGGAGTCTGGGGCGCTGGCGCGCTAGCCGCGC 1500

1501 GCT 1503

1501 GCT 1503

RESULT 2

AAV72539

ID AAV72539 standard; DNA; 1503 BP.

XX AAV72539;

AC AAV72539;

XX 17-OCT-2003 (revised)

DT 01-MAR-1999 (first entry)

XX Thermus flavus amylomaltase encoding DNA.

XX Thermus flavus; amylomaltase; heat resistant; cyclic glucan;

KW intramolecular transglycosylation; alpha-glucan; food; ds.

XX Thermus thermophilus.

PH	Key	Location/Qualifiers	1..1503	301	TGGCTTTACAGACCCGCTGCGGCGGCTTTTCGGGCGGCTTTTCGGGCAAGG	360
FT	CDS	/*tag= a		295	CTCCTCTACGCTGGAAGTGGCCCGCCCTGAGAGAGGCTTTCCGGGGCTTCAGGAAAG	354
FT		/product= "amylomaltase"		361	GCTTCGGCCCAAGGATAAGACCCGACTGGAAGCCCTTTATCGAGGCGGAGCGCTTCTGGCTG	420
XX	EP884384-A2.			355	GCCTCCCGGAGGAGCGGAGGCTTCGCGGCTTCGCGGAGAGGAGGCTGGTGGCTC	414
XX	16-DEC-1998.			421	GAAGACTATGGCTCTTTATGCGCCCTCAAGACCCGCTTTGACGCGAAGCCCTTGAAACGAG	480
XX	13-MAY-1998;	98EP-00250162.		415	GAGGACTACGCTCTCTTATGCGCCCTGAGGGGCGCACGCGGGGCTTCCCTTGAACCGG	474
XX	13-MAY-1997;	97JP-00122635.		481	TGGAGCCCGAGCTGCGGACCGGTAACCGGCTCCCTGCGGAGGCGCGGTGAGAGCTG	540
XX	07-MAY-1998;	98JP-00125121.		475	TGGCCCTTCCCTTGGCGAAGCGGGAAGAGGCTTTAGGGAGGCGGAAGGCGCTTG	534
XX	(EZAK ) EZAKI GLICO CO LTD.			541	GCCGAGGAGTGCGCTTTTACGAGTGGATTCACTGCTTTTATCTGGAGTGGGCGCAG	600
PI	Terada Y, Fujii K, Yanase M, Takata H, Takaha T, Okada S;			535	GCCGAGGAGTGCGCTTCCACGCTTCCACGCTGCTTCTTCGCGCAGTGGGGGCC	594
DR	WPI; 1999-026580/03.			601	ACCAAGCCCTATGCGGATCAAGGGGATTCAGATTATCGGCGATATGCCCATCTTTGTG	660
DR	P-PSDB; RAW83330.			595	TTGAAGCGGAGGCGGAGGCTTGGGCTATCGGATCATCGGGGACATGCCCATCTTCGTG	654
XX	New amylomaltase catalyses intra-molecular trans-glycosylation of alpha-glucans - used as additive to e.g. rice products, snacks, wheat products, noodles, processed seafood, frozen or refrigerated foods, baby foods or drinks.			661	GCCTTCGATTCTCAGATGTCGSGCCAAACCGCAGTACTTCTACCTCGAGGCGGATGCG	720
XX	Claim 5; Page 19; 32pp; English.			655	GCCGAGGACTCCGCGGAGGTCTGGGCGCCACCCGAGTGGTTTACCTTGGACGAGAGGCG	714
XX	The present sequence encodes amylomaltase isolated from <i>Thermus flavus</i> .			721	AACCCCAAGCTGGTGGCGGCGCTTCCGCGGAGTACTTCTCCGAAACCGGCGAGCTCTGG	780
CC	Amylomaltase catalyses intramolecular transglycosylation of alpha-glucans to generate cyclic glucans, has no hydrolase activity, has an optimum temperature of 65-70 degrees Celsius, remains active at 60 degrees Celsius for at least 10 minutes, is inactivated at 100 degrees Celsius within 15 minutes and has an optimum pH of 5.5. Amylomaltase can be used to produce a cyclic glucan by cyclising an alpha-glucan and collecting and purifying the cyclic glucan (especially where the cyclic glucan comprises a cyclic alpha-1,4-glucan, a branched cyclic glucan and/or a branching enzyme is also used in the cyclisation step). The amylomaltase is used to produce food by adding it to a food material before or immediately after cooking so that the amylomaltase acts on starch in the food material to produce a cyclic glucan (especially where the food is selected from rice products, Japanese desserts, snacks, wheat products, CC noodles, gyoza skins, shumai skins, processed seafoods, frozen or CC refrigerated processed foods, weaning foods, baby foods, pet foods, CC animal feeds, drinks, sports foods and nutritional supplements). (Updated CC on 17-OCT-2003 to standardise OS field)			715	CGCCCCAAGCTGGTGGCGGCGGTGCCCCGAGTACTTCTCGGAGACGCGGCGAGCTGG	774
XX				781	GGCAATCCGCTCTATGCTGGGATGATGATAAGGAGCAACTTTGCTGCTGCTGATGCC	840
XX				775	GSTAAACCCCTTTACCGCTGGGACGTTTGGAGCGGAGGCGGTCTCTCTTGTGATCCGC	834
XX				841	CGCATAGGCGAGTCGCTCAAGCAGTGCACCTGCTGGCATCGACCACTTCCGCGGCTTT	900
XX				835	CGTCTGGAGAGGCCCTTGGAGCTTCTCCACCTGCTGGCATAGACCACTTCCGCGGCTTT	894
XX				901	GAAGCCCTACTGGGAGGTTCGTTTGGCGGCGCCCAATGCTGTGGAGGCGCGCTGGGTCAAA	960
XX				895	GAGGCTTACTGGGAGATCCCGCAAGCTGCGCCACGCGGCTGGAGGCGCGCTGGGTCAAG	954
XX				961	GCCCCAGGAGGAGTCTGTTGCTGGGTGCGGCGCCCACTGAGCGATGCGCCCATCTT	1020
XX				955	GCCCCGGGAGGAGTCTTCCAGAAAGATCCAGAGGTCTTTCGCGGAGGTCCCCGCTCTC	1014
XX	Sequence 1503 BP; 223 A; 517 C; 523 G; 240 T; 0 U; 0 Other;			1021	GCCGAGAGCTGGGCGGTGATCACCCCGAGGTGGAGGCTTTTCGCGATGGCTTCGCGGTT	1080
XX	Query Match 42.5%; Score 638.4; DB 2; Length 1503;			1015	GCCGAGAGCTGGGCGGTGATCACCCCGAGGTGGAGGCTTCGCGGAGGCTTCGCGGCTT	1074
XX	Best Local Similarity 65.0%; Pred. No. 1.9e-142;			1081	CCCGGATGAAGATTTTGCAGTTTGTCTTTTTCGCGTGAAGACCAACGCTTTTTCGCC	1137
XX	Matches 977; Conservative 0; Mismatches 516; Indels 9; Gaps 2;			1075	CCCGGATGAAGTCTCGAGTTTCGCTTTGACGACGGGATGAAACCCCTTCTCTCC	1134
QY	1 ATGCAACTCCAAACGCGCTTTTGGAAATTTGCTCCACCCCAACCACTTTTCGCGGTCGCTGG	60		1138	CACAACTACCCCGCGACGCGCAATGTTGGTGTGTACAGCGGAACCCGACGACGACACC	1197
DB	1 ATGAGAGTTTCCCGCGCTTTTCGGTCTGCTCTCCACCCCAAGAGCTTCCCGGCCCCCTAC	60		1135	CACAACTACCTGCCCCACGCGCGGTGTGTGTATACACCGGCAACCCGACGACGACACC	1194
QY	61 GGAATTGGGGCTCTGGGCCCGAGGCGGAGCGGTTTGTGAGCTGGCTGGCGGATGCGGGA	120		1198	ACCCTGGGATGGTTCCGCAACCGCGCGAGGCGCGGCGCTTTCATGCGGGCTACCTG	1257
DB	61 GCGCTGCGGCTCTGGGCGCGGAGGCGCGGAGCTTCTCGCTTCTCTCAAGGAGGCGGG	120		1195	ACCCTGGGCTGGTACCGCAACCGGCAACCCCGGAGAGGAGGCTTTCATGGCGGCTACCTG	1254
QY	121 GCCCGCTGGGAGGTCTTACCGCTGGGCGCTTACCAAGTTACGGCGAGCTGCGGTACAG	180		1258	GCCCGCTATGGATCCGTTTGTTCGGAATAGAGGTTCGCGGCGCTTTTGTATCGAGCTG	1317
DB	121 GGGCGGCTACTGGGAGGTCTTCCCTTGGGCGCCCAACGCGGCTATGGCGACTCCCCCTAC	180		1255	GCGGACTGGGGATCACTTCGCGGAGAGGAGGAGGTGCTTGGGCGCTTGTATGACCTG	1314
QY	181 TCCTTCTCGGCTTTTCGCGGTAAACCGGTATTTGGTTGACCCCGGAGTGTGTAATAA	240		1318	GCCTTCAAAGCCCGGCGCAAGCTGTGTATTTGCTCTTTGAGAGCACTGCTGGGCTGGGC	1377
DB	181 TCCTTACGCGCTTTCGCGCGGAAACCCCTTACTCTATAGACTGAGGCGCCCTCGCGGAAGG	240		1315	GGGATGAAGTCCGTGGCGCGGCTCGCGCTTACCCCGTGTGAGGACGCTCTCTGGCCCTGGC	1374
QY	241 GGTGCTGGAACAAACGAGAGGCGCCCGCGCTATCCGACCCAGGCGGCTGGAATTATGGC	300		1378	CCCGAGGCCCGCATGAACTTCCCGGAGCGCTGGGGGACAACTGGGCGTGGCGCTACGCC	1437
DB	241 GGCTACGTG-----CGCTGGAGGACCCCGGCTTCCCGCAAGGCGGCTGGACTACGCC	294				

||||| 1375 AGCGAGCCCGGATGAACCTACCGGGAAGCCCTCGGGGAACCTGGGCTCGGCTCTCTC 1434  
||||| 1438 GAAGGGGAGCTCGAGCCCGGCTCGGCGGGAAGCTCGGGCCCTCGGAGGCGCAGCCAG 1497  
||||| 1435 CCGGGGAGCTTCCCGGAGCAGCGGGCGAGGCTTAGGGCCATGGCGGAGGCGCAGCGAA 1494  
1498 CG 1499  
1495 CG 1496  
RESULT 3  
ADQ15010  
ID ADQ15010 standard; DNA; 2085 BP.  
XX AC ADQ15010;  
XX AC ADQ15010;  
XX 23-SEP-2004 (first entry)  
XX DE DNA encoding Thermus aquaticus amylo maltase polypeptide.  
XX AMylo maltase; transgenic; foodstuff; food additive; modifier;  
KW saccharide; (alpha)-1,4-glucan; cooked rice; Japanese confectionery;  
KW snack; noodle; Chinese dumpling; shao-mai; fishery kneading;  
KW refrigerated; baby food; pet; animal feed; drink; food supplement;  
KW cyclic glucan; Thermus aquaticus; gene; ds.  
XX Thermus aquaticus.  
XX Key Location/Qualifiers  
FH 514. .2016  
FT CDS /\*tag= a  
FT /\*EC number= "EC 2.4.1.25"  
FT /\*product= "Thermus aquaticus amylo maltase"  
XX JP2004187674-A.  
XX 08-JUL-2004.  
XX 12-SEP-2003; 2003JP-00322319.  
XX 26-NOV-2002; 2002JP-00342966.  
XX (EZAKI) EZAKI GLICO CO LTD.  
XX (NIDE) NEC CORP.  
XX WPI; 2004-528719/51.  
XX P-PSDB; ADQ15011.  
XX New mutated amylo maltase polypeptide having increased enzyme activity  
PT and reduced hydrolyzing activity, useful for manufacturing foodstuffs  
PT such as Japanese confectionery, noodles, baby foods, and food additives.  
XX Claim 5; SEQ ID NO 1; 51pp; Japanese.  
XX The invention relates to a novel amylo maltase polypeptide comprising the  
XX amino acid sequence of a wild-type amylo maltase with a substitution,  
XX addition or deletion at a position, where the amino acid residue  
XX interacts with acarbose of amino acids other than the wild-type amylo  
XX maltase amino acid sequence. The invention further comprises: a nucleic  
XX acid molecule containing a nucleic acid sequence encoding the novel amylo  
XX maltase polypeptide; a vector containing the nucleic acid sequence; a  
XX cell containing the nucleic acid sequence; biological tissue containing  
XX the nucleic acid sequence; a transgenic organism containing the nucleic  
XX acid sequence; a foodstuff, a food additive or a modifier of foodstuff  
XX containing the nucleic acid sequence; a computer readable recording  
XX medium which contains the information of the nucleic acid sequence  
XX encoding the amino acid sequence of the novel amylo maltase polypeptide;  
XX and a cyclic glucan obtained by reacting the novel amylo maltase  
XX polypeptide on saccharides which have a linear structure of (alpha)-1,4-  
XX glucan. The novel amylo maltase polypeptide is useful for manufacturing  
XX foodstuffs such as cooked rice, Japanese confectionery, snacks, bakeries,

CC noodles, Chinese dumpling, shao-mai, fishery kneading goods,  
CC refrigerating foods, baby foods, pet foods, feed for animals, drinks, and  
CC food supplements, where the method involves adding the novel amylo  
CC maltase polypeptide to the foodstuff before or immediately after heat  
CC processing of the raw material, where the novel amylo maltase polypeptide  
CC generates a cyclic glucan from the starch of the foodstuff. The novel  
CC amylo maltase polypeptide is useful for manufacturing glucan which has a  
CC cyclic structure of (alpha)-1,4-glycoside linkage, foodstuff materials  
CC and food additives, modifier of foodstuffs, a drink or eatable  
CC composition, infusion solution or composition for adhesion, where the  
CC method involves reacting the saccharide containing the linear structure  
CC of (alpha)-1,4-glycoside linkage or its derivatives with the novel amylo  
CC maltase polypeptide. This polynucleotide sequence represents the DNA  
CC encoding the Thermus aquaticus amylo maltase polypeptide of the  
XX invention.  
XX SQ Sequence 2085 BP; 318 A; 715 C; 724 G; 328 T; 0 U; 0 Other;  
Query Match 42.5%; Score 638.4; DB 12; Length 2085;  
Best Local Similarity 65.0%; Pred. No. 2.1e-142;  
Matches 977; Conservative 0; Mismatches 516; Indels 9; Gaps 2;  
QY 1 ATGCAACTCCACGCGCTTTTGGAAATTTGTCTCCACCCACCGATTTTCGGGTCTCGTGG 60  
DB 514 ATGGAGCTTCCCGCGCTTTTCGGTCTCTCTCCACCCACGAGCTCTCCCGCCCTAC 573  
QY 61 GGGATGGGGCTCTGGCGCGGAGCGGAGCGGTTTTCGACTGGTGGCCGATCGGGA 120  
DB 574 GCGCTCGGCGTCTCGGCGGAGGCGCGGAGCTTCTCCGCTTCTCAAGGAGGCGGG 633  
QY 121 GCCCGCTGGTGGCAGGTCTTACCGCTGGGCGCTTACCGATTCGCGGACTCGCGTACCAG 180  
DB 634 GGGCGGTACTGGCAGGTCTCTCCCTTGGGCGCCACCGGTATGGGACTCCCGCTACCAG 693  
QY 181 TCCTTCTCGGCTTTTTCGGGTAAACCGTATTTGGTTGACCCCGAGATGCTATTGAAAAA 240  
DB 694 TCTTTCAGGCGCTTCGCGGAAACCCCTACTCATAGACTCGAGGCCCTTCGCGGAAAGG 753  
QY 241 GCGTGGCTGGAAACAAAGCGAGCGGCGCGCTATCCGACCCAGCGCGGTGATATGGC 300  
DB 754 GCGTACGTG-----CGCTGGAGGACCCCGCGCTTCCCGGAGGCGGGTGGACTACGGC 807  
QY 301 TGGCTTTACGACACCGCTGGCGCTTGTTCGCGCGGGCTTTTCGCGGGTTCGCGGCAAGG 360  
DB 808 CTCCTCTACGCTGGAAGTGGCGCGCTTGAAGAGGCGCTTCGCGGGCTTCAGGAAAG 867  
QY 361 GCTTCGCGCCAGGATTAAGACCCGACTGGAAGCTTTATCGAGCCCGAGCGCTTCTGGCTG 420  
DB 868 GCTTCCCGGAGGAGCGGAGGCGCTTCGCGCGCTTCCCGGAGGAGGCGCTGCTGGCTC 927  
QY 421 GAAGACTATGCGCTTTTATGCGCTTCAAGACCCCGTTCAGCGGCAAGCCCTCGMAACGAG 480  
DB 928 GAGGACTACGCGCTTCTCATGCGCTTGAAGGGGGCGCACGGGGGGCTTCCCTCGAACCGG 987  
QY 481 TGGAGCCCGGAGCTGCGCGACCGTGAACCGGCTGCGCTTCCCGGAGGCGCGCTGAGGAGCTG 540  
DB 988 TGCGCTTTCCTTCGCGAGGAGCGGAGAGAGAGGCGCTTAGGGAGGCGGAAAGCGCTTG 1047  
QY 541 GCGGAGGAGTGCGCTTTTACGAGTGGATTCAGTGGCTTTTATCTGGAATGGGCGCAG 600  
DB 1048 GCGGAGGAGTGCGCTTTCACGCGCTTTCAGCGGCTTCTTCTCCGCGGAGGCGGCG 1107  
QY 601 ACCAAGGCTATGCGCAATTCGAAGGGGATTCAGATTCGCGGATGCGGCTATGCGGCTTGTG 660  
DB 1108 TTGAAGCGGAGGCGGAGCGGCTTGGGCGATTCGGGATCATCGGGGACATGCCCATCTTCG 1167  
QY 661 GCCTTCGATTCCTCAGATGTCTGGGCGCAACCCCGAGTACTTCTACCTCGAGGCGGATGGC 720  
DB 1168 GCGGAGGATCTCGCGGAGGTCTGGGCGCCACCCGAGTGGTTTCACTGAGAGGAGGCG 1227  
QY 721 AACCCCGAGGCTGGTGGCGGCGGTTCCGCGGAGTACTTCTCCGAAACCGGCGAGCTCTGG 780  
DB 1228 CGCCCCACGCTGGTGGCGGCGGTTGCCCGGAGTACTTCTTCGGAGAGCGGCGCAGCGCTGG 1287



Qy 301. TGGCTTTACAGACCCGCTGCGCCCTGTTGCGCGCGGCTTTTCGGGGGTTTCGGGCAAG 360  
Db 295 CTCTCTACGCTGGAAGTGGCCCGCTTGAAGAGGCTTTCCGGGGCTTCAAGGAAG 354  
Qy 361 GCTTCGGCCAGGATAAGACCCGACTGAAGCCTTTATCGAGGCGGAGCGCTTCTGGCTG 420  
Db 355 GCCTCCCGAGGAGCGGAGGCTTCCGCCCTTCGGGAGAGGAGGCTTGTGCTC 414  
Qy 421 GAAGACTATGGCTCTTTATGGCCCTCAAGACCCGGTTTGAAGCAAGCCTTGAAGAG 480  
Db 415 GAGGACTACGCCCTCTTATGCGCCTGAAGGGGCGCACGGGGGCTTCCCTGGAACCG 474  
Qy 481 TGGAGCCCGAGCTGCGGACCGTGAACCGGCTGCTGCGCCAGGCGCCGCTGAGAGCTG 540  
Db 475 TGGCCCTTCCCTTCCGGAAGCGGAGAGAGAGGCGCTTAGGGAGCGGAAGCGCCTTG 534  
Qy 541 GCCGAGGAGTGGCCCTTTACGAGTGGATTACAGTGGCTTTTATCTGGAATGGGGCCAG 600  
Db 535 GCCGAGGAGTGGCTTCCAGCCTTACAGTGGCTTCTTCGCGCAGTGGGGGCC 594  
Qy 601 ACCAAGCCTATCGGAATCAAGGGGATTAGATTATCGCGGATATGCCCATCTTTGTG 660  
Db 595 TTGAAGCGGAGGCGAGCGCTTGGCATCCGATCATCGGGACATGCCCATCTTCGTG 654  
Qy 661 GCCTTCGATTCAGATGCTGGGCCAAACCGGAGTACTTCTACCTCGAGCGCGATGCG 720  
Db 655 GCCGAGGACTCCGCGAGTCTGGGCCCAACCGGAGTGGTTTCACTTCGAGGAGGAGG 714  
Qy 721 AACCCGAGTGGGGGGGTTCCGGGAGTACTTCTCGAAACCGGCGAGCTCTGG 780  
Db 715 CGCCCGAGTGGTGGGGGGTGGCCCGGACTACTTCTCGGAGACGGGCGAGGCTGG 774  
Qy 781 GCGAATCGCTCTATCGCTGGGATGTGATGAAAGGGAACAATTTCCTGCTGTGATGCC 840  
Db 775 GCGAACCCCTTACCGCTGGGACGTTTGGAGCGGAGGGGTTCTCTCTGATCGC 834  
Qy 841 CGCATAGGAGTCTCAGCAGTGCACCTGTGCGCATCGACCTTCGCGGGTTT 900  
Db 835 CGTCTGGAAGAGGCGCTCGAGCTCTCCACCTGTGCGCATAGACCACTTCGCGGGCTT 894  
Qy 901 GAGCCTACTGGAGTTCGTTTGGCGCGCCATGCTGTGGAGGGCGCTGGGTCAA 960  
Db 895 GAGGCTACTGGAGATCCCGCAAGCTGCCACCGGCTGGAGGGCGCTGGGTCAAG 954  
Qy 961 GCCCGAGGAGAGCTGTTGCTCGGTCGGGCGCCAACTGAGCGATGCGCCCATATT 1020  
Db 955 GCCCGGGGAGAGCTCTTCAGAGATCCAGGAGGTTTCGGGAGGTCCCGCTCCTC 1014  
Qy 1021 GCGGAAGACTTGGGGGTATCACCCCGAGGTGGAGGCTTTCGGCGATGGCTTCGGGTC 1080  
Db 1015 GCGGAGGACTTGGGGGTATCACCCCGGAGGTGGAGGCGCTTCGCGACCGCTTCGGGCTT 1074  
Qy 1081 CCCGCGATGAGATTTGAGTTCGTTTTCGGAATACGAGTTCGCGGCGCTTTTTCGCC 1137  
Db 1075 CCCGCGATGAGGTCCTGAGTTGCGCTTTGACGCGGATGGAACCCCTTCTCCTCC 1134  
Qy 1138 CACAACTACCCCGCAGCGCAATGTGTGTGTATACGCGGAACCCACGACAAACACACC 1197  
Db 1135 CACAACTACCCCTGCCACCGCGGGTGTGTGTATACCGGCAACCCACGACAAACACACC 1194  
Qy 1198 ACCCTGGAGTGTTCGCAACCGCGCGGAGGCGGCGCTTCATGCGGCGCTACCTG 1257  
Db 1195 ACCCTGGGCTGTACCGCAGCGCCACCCCGCAGAGAGGCGCTTCATGCGGCGTACCTG 1254  
Qy 1258 GCCCGCTATGGCATCGTTTGTTCGGAATACGAGTTCGCGGCGCTTTGATCGAGCTG 1317  
Db 1255 GCGGACTGGGGATCACCTTCGGAAGAGGAGGAGGCTTCCTGGGCGCTGATGACCTG 1314  
Qy 1318 GCTTCAAAGCCCGCAGCTGGCTATGTGCTTTGACGAGCTGCTGGGCTGGG 1377  
Db 1315 GGGATGAGTCTGTCGGCGGCTCGCGTCTACCCCGGTGACGAGCTGCTGGCCCTGGG 1374  
Qy 1378 CCCGAGGCGCGCATGAATCTTCCCGGACGCTTGGGGGACAACTGGGCGTGGCGCTACGCC 1437

Db 1375 AGCGAGCCCGGATGAACTACCGGAGGCGCTTCGGGAACTGGGCTGCGGCTCTC 1434  
Qy 1438 GAAGCGGACCTCGAGCCGCTTGGCCCGCGGACTGCGGCGCTTCGGCGGAGCCAGCAG 1497  
Db 1435 CCGGGGAGCTTTTCCCGGAGCACGGGCGAGGCTTAGGGCCATGGCGGAGGCCACGAA 1494  
Qy 1498 CG 1499  
Db 1495 CG 1496

RESULT 5  
ADU07728  
ID ADU07728 standard; DNA; 1539 BP.  
XX  
AC ADU07728;  
XX  
DT 13-JAN-2005 (first entry)  
XX  
DE DNA sequence #221 encoding amylase.  
XX  
KW Amylase; glucoamylase; 1,4-alpha-D-glucan glycohydrolase; alpha-amylase;  
KW exoamylase; beta-amylase; glucosidic bond hydrolysis; starch; sugar;  
KW glucose; maltodextrin; thermostable; gene; ds.  
XX  
OS Unidentified.  
XX  
PN W02004091544-A2.  
XX  
PD 28-OCT-2004.  
XX  
PF 08-MAR-2004; 2004WO-US007096.  
XX  
PR 06-MAR-2003; 2003US-00385305.  
PR 28-MAR-2003; 2003US-0459014P.  
XX  
PR (DIVE-) DIVERSA CORP.  
XX  
PI Callen W, Richardson T, Frey G, Gray K, Kerovuo JS, Slupska M;  
PI Barton N, O'donoghue E, Miller C;  
XX  
XX WPI: 2004-775532/76.  
XX P-PSDB; ADU07729.  
XX  
XX Novel recombinant amylases, useful for hydrolysis of starch into sugars,  
XX in detergent composition, for producing high-maltose or high-glucose  
XX syrup, in brewing or alcohol production, bio-bleaching composition of  
XX paper or pulp product.  
XX  
XX Claim 4; SEQ ID NO 438; 436pp; English.  
XX

The invention relates to the isolation of polypeptides with amylase activity, and polynucleotide sequences encoding the polypeptides. The amylase activity comprises a glucoamylase activity, 1,4-alpha-D-glucan glycohydrolase activity, alpha-amylase activity, exoamylase activity or beta-amylase activity. The amylase activity comprises hydrolyzing glucosidic bonds in a starch to produce sugars such as glucose and maltodextrins. The amylase activity comprises cleaving a maltose or a D-glucose unit from the non-reducing end of the starch. The amylase activity is thermostable and the polypeptides of the invention are useful over a wide range of temperatures. The polypeptide sequences are useful for making anti-amylase antibodies, for identifying a modulator of an amylase activity, or identifying an amylase substrate. The polynucleotide sequences are useful for producing recombinant polypeptides, and for modifying codons in a nucleic acid encoding a polypeptide with an amylase activity to increase or decrease its expression in a host cell. The polypeptide and polynucleotide sequences are useful for making or modifying a small molecule, and for determining a functional fragment of an amylase enzyme. Primers to the polynucleotides are useful for amplifying them. This sequence represents a polynucleotide sequence of the invention.



SQL	Sequence	1539 BP; 385 A; 401 C; 378 G; 375 T; 0 U; 0 Other;
Query Match	25.1%;	Score 378; DB 13; Length 1539;
Best Local Similarity	55.7%;	Pred. No. 3.3e-80;
Matches	801; Conservative	0; Mismatches 610; Indels 27; Gaps 3;
QY	1	ATGCAATCCAAACGGCTTTTGGAAATTTTGTCCACCCACAGTTTTCGGGTGCTGG 60
DB	1	ATGAAATTAAGCGGCTCGGGAATCATCTTATCTACAGCTTCCGGGCGCGAT 60
QY	61	GGATTGGGCTCTGGGCGGAGCCGAGCGGTTTTGGACTGGCTGGCCGATCGGGA 120
DB	61	GGAAATGGTGACCTTGGACCGGAAGCATACGCTGGATCAACTTTCTAGCCGAATCCGGC 120
QY	121	CCCGCTGGTGCGAGTCTTACCGCTGGGCGCTTACAGTTACGGCGACTCGCCGTACACG 180
DB	121	TGCAAACTATGGCAGTCTCGCGCTTGGCCCAACCGGTACGGCGACTCGCCCTATCAG 180
QY	181	TCCTTCTCGGCTTTTGGCGGTAAACCGGTATTTGGTTGACCCCGAGATGCTGATTTGAAAAA 240
DB	181	TGTTTCTCAGCATTTGCGCGTAAATCCCTATCTGCTACGCCCGCGCTGTTGCTGGAAC 240
QY	241	GGCTGGTGGAACAAAGCGA-----AGCGCCCGCGCGTATCCGACCCAGCGGTGGAT 294
DB	241	GGTTTGTACGGCGGAGTGATCTGACCGACCGCCCTGAAATTTAATCCAAATTCGAATTTGAT 300
QY	295	TATGCTGGCTTTACAGACCGCTGGCGCCCTGTTTGGCGGGCTTTTCGGGGGTTTCGG 354
DB	301	TACGGTGAAGCCATCATTGGAATTAATAAATACTCGATCGCGCTTACGAAAACTTTAAG 360
QY	355	GCAAGGGCTTCGGCCCGAGGATAGACCCGACTGGAAGCCTTTATCGAGGCGCGCGCTTC 414
DB	361	CGCTCAATATCTGAACTAAATATGAATATCAGCAATTCACGGAATATGACGAGAC 420
QY	415	TGGCTGGAAGACTATGGCTCTTTATGGCCCTCAAGACCCCGTTTGAAGCAAGCCCTGG 474
DB	421	TGGTTGAATGATTTTGGCCCTTTTATGGCCATTAAGGAGGCTAACGGGGGGGTATCCCTG 480
QY	475	AACGAGTGGAGCCCGGAGCTCGCGACCGTGAACCGGCTGCTGCGCGAGGCGCGTGAG 534
DB	481	GACAACTGGCCAAAGAACTGCGTAGCGCCGCGAGCGCAATCGAAAAAGTTCAAACAA 540
QY	535	GAGCTGCGCGAGGAGTGGCGCTTTTACGAGTGAATTCAGTGGCTTTTATCTGGAATGG 594
DB	541	ACCGAAGCCGACCGCATTAACGGCCATGCTTTCGACAGTTTATTTCTTCGCGCATGG 600
QY	595	GGCCAGAACCCAGGCTATGCGCAATCCAAAGGGATTCAGATATTCGCGCATATGCCCATC 654
DB	601	CTTGACCTGAAAGCCTACGCGCAATCAAAAGAAACATCAAAATCATCGGTGATATACCGATT 660
QY	655	TTTGTGGCTTCCGATTCCTCAGATGCTGGGCGCAACCGGAGTACTTCTACCTCGAGCC 714
DB	661	TTTGTGCTCTACGATACGCGAGATGCTTGTGCGAACCACAGTGTGTTTATCTGGAATGAA 720
QY	715	GATGGCAACCCACGCTGGTGGGCGGCTTCGCGGAGTACTTCTCCGAAACCGGCGAG 774
DB	721	GAAGGCAACCAACTGTTGGCGAGGCTTCCACCGGATTTATTTTCCCAACGGGTCAA 780
QY	775	CTCTGGGCAATCCGCTCTATCGCTGGGATGTGATGAAAGGAGCAACTTTTGGCTGGTGC 834
DB	781	TTATGGGGTAACCGCTTTTACAAGTGGGAAGTTTCATCGTCAGCAGAAATTTCCGCTGGTG 840
QY	835	ATTGCCCGAATAGGAGTCTCAAGCAGTGCACCTGCTGGCATCGACCACTTCGCG 894
DB	841	ATTCCGCGGATGCAAGCCACATACCGCTGGTATGATTTGTCCGCTCGACCAATTTTCGC 900
QY	895	GGGTTTGAAGCTTACTGGGAGGTTTCGTTTGGCGGCGCAATGCTGTGGAGGCGCGCTGG 954
DB	901	GGATTTGCCGGTATTTGGGAAGTACCTTTTGGATGCCACCGCGGAAATTTGCGCGCTGG 960
QY	955	GTCAAGCCCCAGGGAGAGCTGTTTGTGCTGGGTGGCGGCCCAACTGAGCGATGCGGCC 1014
DB	961	GTAAGGGGCGCGCAAGAAATTTATCAATGCCATTTCGGGATGCGCTTGGCGATTTACCC 1020

QY	1015	ATCATTCGCAAGACCTGGGGTGATCACCCCGAGGTGAGGCTTTTCGCGCATGGCTTC 1074
DB	1021	ATCATCGCTGAAGATTTGGGGTCAATCACCCCGATGTGATTGAATGCGGATTCGTTT 1080
QY	1075	GGGTCCCGCGCATGAAGATTTTTCAGTTTTCGGTTCGGGTGAGGCAACGCTTTTGG 1134
DB	1081	AACCTGGCCGCTATGAAGGTGTTCCAGTTTGTCTTTTACAACGAGCCCGCTCGATCCCTTC 1140
QY	1135	CCCCACAACTACCCCGCGCAGCGCAATGTGGTGTACAGCGAAACCCACGACAAACGAC 1194
DB	1141	CTGCGCGCAACAACACTAGC---AAACCAATTTGTGTAGCTTACACTGTTACCCAGCAACGAC 1197
QY	1195	ACCACCTTGGGATGTTTCCGCAACCGCGCGAGCGCGGCTTTCATCGCGGCGCTAC 1254
DB	1198	ACGCTTGGGTGGTATCAAGGCGCACCGGAAAGAAAGACGTGATTTTATTCGCGGATAT 1257
QY	1255	CTGCGCGCTATGGCATFCCGTTTGTTCGGAATACGAGGTTCGCGGCGCTTTGATCGAG 1314
DB	1258	CTGCGCGTTTCCGCGC-----GAAGATATATCATGGGACATGATCGC 1299
QY	1315	CTGCGCTTCAAAAGCCCGCAAGCTGGCTATTTGCTTTCGAGGAGCTGCTGGGCTG 1374
DB	1300	GCTGTGTGTCTTTCAGTGGCGCTTTTTCGCTTTCGCGCCATGCAAGATATTTTCAGCGCTT 1359
QY	1375	GGCCCGGAGCGCGCATGAATCTTCCCGGACGCGCTGGGGGACAACTCGGCGTGGCGCT 1432
DB	1360	GGAATGAAGCCCGCATGAATCTTACCGGGAAGACCGAGCGGCACTGTTGTGGAGAT 1417

RESULT 6  
ADQ15025  
ID ADQ15025 standard; DNA; 1518 BP.  
XX AC ADQ15025;  
XX DT 23-SEP-2004 (first entry)  
XX DNA encoding Synechocystis sp. PCC 6803 amylo maltase polypeptide.  
XX Amylo maltase; transgenic; foodstuff; food additive; modifier;  
XX saccharide; (alpha)-1,4-glucan; cooked rice; Japanese confectionery;  
XX snack; noodle; Chinese dumpling; shao-mai; fishery kneading;  
XX refrigerated; baby food; pet; animal feed; drink; food supplement;  
XX cyclic glucan; Synechocystis sp. PCC 6803; gene; ds.  
XX Synechocystis sp. PCC 6803.  
XX Key Location/Qualifiers  
XX CDS 1. .1518  
XX /tag= a  
XX /product= "Synechocystis sp. PCC 6803 amylo maltase  
XX polypeptide"  
XX JP2004187674-A.  
XX 08-JUL-2004.  
XX 12-SEP-2003; 2003JP-00322319.  
XX PF P-PSDB; ADQ15026.  
XX PR 26-NOV-2002; 2002JP-00342966.  
XX (EZAKI) EZAKI GLICO CO LTD.  
XX (NIDE) NEC CORP.  
XX WPI; 2004-528719/51.  
XX P-PSDB; ADQ15026.  
XX New mutated amylo maltase polypeptide having increased enzyme activity  
XX and reduced hydrolyzing activity, useful for manufacturing foodstuffs  
XX such as Japanese confectionery, noodles, baby foods, and food additives.  
XX Example 9; SEQ ID NO 16; 51pp; Japanese.

xx CC The invention relates to a novel amylo maltase polypeptide comprising the  
 CC amino acid sequence of a wild-type amylo maltase with a substitution,  
 CC addition or deletion at a position, where the amino acid residue  
 CC interacts with acarbose of amino acids other than the wild-type amylo  
 CC maltase amino acid sequence. The invention further comprises: a nucleic  
 CC acid molecule containing a nucleic acid sequence encoding the novel amylo  
 CC maltase polypeptide; a vector containing the nucleic acid sequence; a  
 CC cell containing the nucleic acid sequence; biological tissue containing  
 CC the nucleic acid sequence; a transgenic organism containing the nucleic  
 CC acid sequence; a foodstuff, a food additive or a modifier of foodstuff  
 CC containing the nucleic acid sequence; a computer readable recording  
 CC medium which contains the information of the nucleic acid sequence  
 CC encoding the amino acid sequence of the novel amylo maltase polypeptide;  
 CC and a cyclic glucan obtained by reacting the novel amylo maltase  
 CC polypeptide on saccharides which have a linear structure of (alpha)-1,4-  
 CC glucan. The novel amylo maltase polypeptide is useful for manufacturing  
 CC foodstuffs such as cooked rice, Japanese confectionery, snacks, bakeries,  
 CC noodles, Chinese dumpling, shao-mai, fishery kneading goods, and  
 CC refrigerating foods, baby foods, pet foods, feed for animals, drinks, and  
 CC food supplements, where the method involves adding the novel amylo  
 CC maltase polypeptide to the foodstuff before or immediately after heat  
 CC processing of the raw material, where the novel amylo maltase polypeptide  
 CC generates a cyclic glucan from the starch of the foodstuff. The novel  
 CC amylo maltase polypeptide is useful for manufacturing glucan which has a  
 CC cyclic structure of (alpha)-1,4-glycoside linkage, foodstuff materials  
 CC and food additives, modifier of foodstuffs, a drink or eatable  
 CC composition, infusion solution or composition for adhesion, where the  
 CC method involves reacting the saccharide containing the linear structure  
 CC of (alpha)-1,4-glycoside linkage or its derivatives with the novel amylo  
 CC maltase polypeptide. This polynucleotide sequence represents the DNA  
 CC encoding a *Synechocystis* sp. PCC 6803 amylo maltase polypeptide of the  
 CC invention.

xx SQ Sequence 1518 BP; 362 A; 392 C; 390 G; 374 T; 0 U; 0 Other;

Query Match 21.9%; Score 329.2; DB 12; Length 1518;  
 Best Local Similarity 53.3%; Pred. No. 1.5e-68;  
 Matches 800; Conservative 0; Mismatches 673; Indels 27; Gaps 4;

QY 1 ATGCACTCCACCGCTTTGGATTTTGGCTCCACCCACCGAGTTTTCGGGTGCGTGG 60  
 DB 1 ATGTTAGATAAACCGCTGAGCGGTATTCTGCTCCATCCACCTCTCTGCCAGTCGTTTC 60  
 QY 61 GGGATTGGGCTCTGGGCGCGAGCGCGGTTTGGACTGGCTGGCGCGATGCGGGA 120  
 DB 61 GGCATTGGGACTTGGGGAGCGGTCTTTTTCAGTTCATCGATTTTTGGCGAGCGAGAC 120  
 QY 121 GCCGCTGGTGGCAGGTCTTACCGCTGGGCGCTACCGATTACGGGAGCTCCCGTACGAG 180  
 DB 121 CAGAGTGATGGCAATATTGCGGTTGGGCGCCACCGGATTTCGGCAATTCCCTTACCTT 180  
 QY 181 TCCTTCTGGCTTTTGGCGGTAAACCGATTGTTGGTTGACCCCGAGATGCTGATTGAAAAA 240  
 DB 181 TGCTATTCTGCGCTAGCAATTAATTCCTTGGTGTATTGCTTGTATGCTGCTAGCTGAAGA 240  
 QY 241 GGCTGGCTGGAACAAAG-----CGAAGCGGCGCGCGCTATCCGACCCAGCGGTGGAT 294  
 DB 241 GGATTTTGGCAGGATTATTAGCAAGCGCGCGCGCTTACTTAACCCAGGCTGGAT 300  
 QY 295 TATGCTGGCTTTTACAGACCGCTGGCGCTGTTGGCGGCGGCTTTCGGGGTTTGG 354  
 DB 301 TATGACCAAGCGGATGCGCTACAAATCCCAAGTTTAAACACAGCGCTTTGCGCCAGTTCCGT 360  
 QY 355 GCAAGGCTTGGCGCCAGGATNAGACCGCTGGAGCTTTATCGAGCGGAGCGCTTC 414  
 DB 361 ACCAATAGAGCTAGCAATTGAGCGAATTTGCGAATTTTGCAGACCCCAAGTAT 420  
 QY 415 TGGCTGGAGACTATGCGCTCTTTATGGCCCTCAAGACCCGCTTTGACGGCAAGCCCTGG 474  
 DB 421 TGGCTAGCAGATTACCGCTGTTTATGGCCATCAAGAGCCCAATGAGCGGCTGG 480  
 QY 475 AACGAGTGGAGCCCGCGAGCTGGCGAACCGGTGAACCGGCTGCGCTGGCGAGGCGCGTGAG 534

DB 481 CATCAATGGACNAGAGCAATTGCTGGCGGGAACCGGAAGCCCTGAAATCTCGGGCGAT 540  
 QY 535 GAGCTGGCGGAGAGGTGGCCCTTTACGAGTGAATTCAGTGGCTTTTATCTGGATGG 594  
 DB 541 CGCCTGAAACCGAAGTTTTATACCATCAATTTTGCATTTCTGGTTTTCGCAATGG 600  
 QY 595 GGCACAGCAAGCCCTATCCGATCCAGGGAATTCAGATTATCGCGATATGCCCATC 654  
 DB 601 CAGGAAGTCAAGCCCTACGCTAACCAACGCGCATAGCCATTTTTGGGAGCTACCCATC 660  
 QY 655 TTGTGCGCTTCCGATTCCTCAGATGCTGGGCGCAACCCGCGAGTACTTCTACCTCGAGGCC 714  
 DB 661 TAGTTGCCACAGTAGCGCTGACGTTTGGGCCAATCCAGAGAACTTTTGCCTTGATCCG 720  
 QY 715 GA---TGGCAACCCCAACGGTGGTGGCGGGGTTTCCGGGAGCTACTTCTCGAAGACCGGC 771  
 DB 721 GAAACCGGAGAGCGGCATGATGGCGGGGTACCAACCGGACTATTTCAGTGGCACAGGG 780  
 QY 772 CAGCTCTGGGCAATCGCTCTATCGTGGATGTGATGAAAGGACAACTTTGCGCTGG 831  
 DB 781 CAACTCTGGGGTAACTCTGTGACGACTGGGAAACCTCNAAGCTACGGGCTTTGCTGG 840  
 QY 832 TGCATTGCGCGCATAAAGCGAGTGGCTCAAGCAGTGGCACCTGGTGGCGCATCGACATTC 891  
 DB 841 TGGATTAAAGCGTTTAAAGCCAATCTGCAATATCTAGACATTGTCGCAATTGACCATTC 900  
 QY 892 CGCGGTTTGAAGCCTTACTGGGAGGTTTCGTTTGGCGGCGCAATGCTGTGGAGGGCGC 951  
 DB 901 CGGGGCTTCGAGTCTTATTTGGGAGTGGCCCAAGGGGGAATAAACTGCTGAAAAATGGCGAA 960  
 QY 952 TGGGTCAAGCCCGGAGGAGAGCTGTTGCTGGTGGCGGCCCAACTGAGCG---AT 1008  
 DB 961 TGATATCCAGCCCCCGGCAAGGAATTTTCCAAGCCTAGGGAAGCCCTGGGTGATAAT 1020  
 QY 1009 GCGCCCATCATTTGCCGAAGACCTGGGGGTGATCACTCCCGAGGTGGAGGTTTTCGCGAT 1068  
 DB 1021 TTACCCATTTGCGCGGAGATTTGGGGGTAAATTTACTCCGGAAGTGAAGCGCTACGGAT 1080  
 QY 1069 GGTTCGGGTTCGCGCATGAAGATTTTTCAGTGGTCTTTTCCGGTGGAGGCAACGCC 1128  
 DB 1081 GAATTTAACTTTCCCGCATGAAAGTGTCTCATTTGCGCTTTGACTCCGACCGGGGTAAT 1140  
 QY 1129 TTTTGGCCCACTACTACCCCGCGCAGCAATGCTGTGTGTACAGCGGAACCCAGAC 1188  
 DB 1141 CCCTTTCTGCGCTTCAACTACAGCAATGGCAACCGGTGTATATCCGCGACCCAGAC 1200  
 QY 1189 AACGACACCACTCGGATGTTCCGCAACCGCGCGGAGCGCGGCTTTCATGCGG 1248  
 DB 1201 AACGACACCACTCGGCTGGTTCAGGAACGGTTCAGAGGATGATCAGC----- 1249  
 QY 1249 GCCTACCTGCGCGCTATGGCATCGGTGTTTGGGAATACGAGGTTCGCGGGCGCTTG 1308  
 DB 1250 -----AAAAGGTGATCAATTACTCGGCTGTGTGCAACGAAGGATATCCATTGAGCTTA 1305  
 QY 1309 ATCGAGTGGCTTCAAAAGCCCGGCCAAGCTGCTATTGTCCTTTGAGGAGCTGCTG 1368  
 DB 1306 ATTGCGCTAGCGTCAAGTTCGGTGGCGGCGCTAGCAATTTTCCCTCCAGGATCTCTC 1365  
 QY 1369 GGGTGGGCGCCGAGGCGCCCATGAACTTCCCGCGAGCGGTGGGGGCAACTGGCGGTGG 1428  
 DB 1366 GGTGGGTAGTGNCTGTCGATGAATTAACCGGCACTGCGGCGGCAACTGGGGTTGG 1425  
 QY 1429 CGCTACCGCAAGGCGACCTGAGCGCGGCTGCGCGGAGCTGCGGGCGCTTGGCCGAG 1488  
 DB 1426 GTTTACCATCCCGACCAACTCAACGATTGGCTATCGGGGCAATTTGAGTTTCATTACGGAG 1485

RESULT 7  
 ADK45052  
 ID ADK45052 standard; DNA; 1518 BP.  
 XX  
 AC ADK45052;

XX 24-FEB-2005 (first entry)  
 XX Streptococcus pneumoniae gene, Seq ID No 1567.  
 XX ds; gene; Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.  
 XX Streptococcus pneumoniae.  
 XX US6699703-B1.  
 XX 02-MAR-2004.  
 XX 26-MAY-2000; 2000US-00583110.  
 XX 02-JUL-1997; 97US-0051553P.  
 XX 12-MAY-1998; 98US-0085131P.  
 XX 30-JUN-1998; 98US-00107433.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 XX Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;  
 XX WPI; 2004-212399/20.  
 XX P-PSDB; ADK47713.  
 XX New nucleic acid molecules and polypeptides useful for diagnosing,  
 XX preventing and treating pathological conditions resulting from bacterial  
 XX infection, e.g. Streptococcus pneumoniae infection, and in drug  
 XX screening.  
 XX Disclosure; SEQ ID NO 1567; 301pp; English.  
 XX The invention relates to isolated Streptococcus pneumoniae nucleic acids  
 XX and polypeptides. The nucleic acids and proteins are useful for  
 XX diagnosing, preventing and treating pathological conditions resulting  
 XX from bacterial infection, such as S. pneumoniae infection. These may also  
 XX be used for drug screening procedures. The present sequence represents a  
 XX Streptococcus pneumoniae nucleic acid of the invention. Note: The  
 XX sequence data for this patent did not appear in the printed specification  
 XX but was obtained in electronic format directly from USPTO at  
 XX seqdata.uspto.gov/sequence.html.  
 XX Sequence 1518 BP; 435 A; 308 C; 351 G; 424 T; 0 U; 0 Other;  
 Query Match 19.2%; Score 289; DB 13; Length 1518;  
 Best Local Similarity 51.7%; Pred. No. 6.2e-59;  
 Matches 764; Conservative 0; Mismatches 695; Indels 18; Gaps 4;  
 11 AACCGCTTTTGGAAATTTGCTCCACCCACCACCTTTTCCGGTCCGCTGGGGATTGGG 70  
 8 AACGTCAAAGTGGTGTGATGACATCTCTTCTCCAGGAGCTTACGGAATCGAT 67  
 71 CTCGCGCGCGAGCGCGGCTTTTGGACTGGCTGGCGGATGCGGAGCCCGCTGGT 130  
 68 CATTGGTCAAAGTCTTACGACTTGGTGGTTCCTTGGTCCGTACAAACAGTTACT 127  
 131 GGCAGGTCTTACCGCTGGGCCCTACAGTTACGCGACTCGCCGTACCGTCTTCGG 190  
 128 GGCATACTCTCCATTAGGAGCACTAGTTACGGGATTTCTCTTACCAATCTTCTCAG 187  
 191 CTTTGGCGGTAAACCGTATTTGGTTGACCCCGAGATGCTGATGAAAAGGCTGGCTGG 250  
 188 CTTTCGAGGAAACACTCAATTTATGATTTAGATATCTTGGTGGAGCAAGGTTTGTGG 247  
 251 AACAAAGCGAAGCGCCCGCGGTATCCGACCCAGCGGTGATTTAGTGGCTGGCTTTACC 310  
 248 AAGCAAGTACCTTGAAGGAGTTGACTTTGGTAGCGATGGCTGGAAGTTGACTATGCTA 307  
 311 AGACCCCGCTGGCCCTGTGTGGCGGGCTTTTCGGGGGTTTCGGGCAAGGCTTCGGCCC 370  
 308 AATCTACTATGACGTCGTCCTCTTTTAGAAAAGCGGTGAACGTTTCTTTGAAAGTCG 367  
 371 AGGATAAGACCCGACTGGAAGCCCTTTATCGAGGCGGAGCGCTTCTGGCTGGAAGACTATG 430  
 368 GAGATGTTAAAGATTTTGAAGAAATTTGCTCAAGACAACCAATCATGGCTTGAGCTTTTG 427  
 431 CGCTCTTTATGGCCCTCAAGACCCGGTTTGAAGCAAGCCCTGGAAACAGTGG---AGCC 487  
 428 CTGAGTATATGGCTATCAAGAGCAATTTGACAAATCTTGTGTGAGCTGGAATGGCCAGATG 487  
 488 CCGAGCTGCGCGACCGGTGAACCGGCTCCCTGGCCAGGCGGCTGAGAGCTGGCCGAGG 547  
 488 CAGATGCTGCTGCTGTAAGCTTCAGACTTGAAGCTATCGTAGCAATTTGGCAGACA 547  
 548 AGGTGGCCCTTTACGAGTGGATTTCAGTGGCTTTTATCTGGAATGGGGCCAGACCAAGG 607  
 548 AGTTGGTTTACCACCGGTGACTCAATACTTCTTCCAAACATGTTGAAATTTGAAG 607  
 608 CTTATGCCGAATCAAGGGGATTCAGATTATCGGCGATATGCCCATCTTTGTGGCTTCG 667  
 608 CTTACGCTAACGACCAACCATCGAATCGTTGGGACATGCCAATCTACGTAGCGGAAG 667  
 668 ATTCTCAGATGCTGGGCCAACCGCGAGTACTTCTACTCTCGAGGCGGATGGCAACCCA 727  
 668 ATTCAAGTGATATGTGGGCAATCCACATCTCTTCAAAACAGATGTCAATGGTAAGGCTA 727  
 728 CGGTGGTGGCGGCTTCCGCGGACTACTTCTCCGAAAACCGCCAGCTCTGGGGCAATC 787  
 728 CTTGTATCGCAGATGCCACACAGATGAGTTTCTGTAACTGGTCAGCTTTGGGGTAATC 787  
 788 CGCTCTATCGCTGGATGTGATGAAAGGACAACTTTTGGCTGTGTGATTCGCCGCAATA 847  
 788 CAATCTATGACTGGGAAGCAATGGAACAAAGCGCTACAAATGGTGGATTGAACGCTTGC 847  
 848 GGCAGTCCGCTCAAGCAGTGCCACTGTGTGCGATCGACCACTTCCGGGGTTTCAAGCCT 907  
 848 GTCAAGCTTCAAAATCTACGATATCGTTCTGTATCGACCACTTCCGGGGCTTCAATCTT 907  
 908 ACTGGGAGGTTTCCGCTTTGGCGGCGCCCAATCTGTGGAGGGCGCTGGGTCAAAAGCCGAG 967  
 908 ACTGGGNAATCCCTGCTGGTTCGATACAGCAGCACTGTGTGGTGGTGAAGSTCCAG 967  
 968 GGGAGAGCTGTTGTGCTGGTGGCGGCCCAACTGAGCGATGCGCCCATCATTTGCCGAG 1027  
 968 GCTACAAGCTTTTTCAGCGCGTTAAAGGAAGAACTTGGTGAGCTAAACATCATCGCAGAG 1027  
 1028 ACTGGGGTGATCACCCCGAGTGGAGCTTTGGCGATGGCTTGGGTTCGCCGCA 1087  
 1028 ACTTGGCTTTCATGACAGATGAAAGTGAATGCGAATTTGGGTGAACGTAAGCTTCCAGAA 1087  
 1088 TGAAGATTTTGCAGTTTGTCTTTTCCGGTGAGCAACAGCGCTTTTTCGCCCAACACTACC 1147  
 1088 TGAAGATTTTCAATTTGCTTCAACCCAGAAAGCGAAGCATTTGATAGCCCACT--- 1144  
 1148 CCGCGCACGCAATGTGTGGTGTATACAGCGGAACCCACGAAACGACACACCCCTGGAT 1207  
 1145 TGGCACCTGTAACTCAGTTATGTACACAGGAACACACGATAACATACTGCTTCTGGTT 1204  
 1208 GGTTCGCGACCGCGCGAGGCGCGAGGGCGCTTTCATGCGGGCTTACTGGCCGCTATG 1267  
 1205 GGTACCGTATGAGATTTGATGATGCG-----ACTCGTGTGATACATGGCTGTTACA 1255  
 1268 GCATCGTTGTTTGTGCGAATACAGGTTCGCGGCGCTTTTGTATCGAGCTGCGCTTCAAAA 1327  
 1256 CGAACCGTAAAGATACGAAACAGTGGT---ACACGCTATGCTTCGTACAGTATTTTCAT 1312  
 1328 GCCCGGCCAAGCTGGCTATTTGCTTTTCAGGACGTTGTGGGGCTGGGCCCGAGGCC 1387  
 1313 CAGTGTAGCTTTATGGCAATTTGCAACTATGCAAGATTTACTAGAAATGGATGAGGAGCTC 1372  
 1388 GATGAATCTTCCCGGACGCTGGGGGACAACTGGGGGTGGCGCTAGCCCGGAAGGGACC 1447  
 1373 GTATGAATCTTCCCATCTACCTTGTGGTGAACCTGGTCTTGGCGTATGACTGAAGATCAAT 1432  
 1448 TCAGAGCCCGTCTCTGGCGCGGGGACTGCGGGCCCTGGC 1484

Db 1433 TGACACCAGCTGTCGGAGGAAGTTTGCTTGCATCGAC 1469

RESULT 8  
ABX07913 standard; DNA; 1515 BP.

XX ABX07913;  
AC  
XX 27-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)

XX S. pneumoniae type 4 strain coding region #2201.  
DE  
XX Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;  
KW ear infection; antinflammatory; antibacterial; immunostimulant;  
KW auditory; respiratory; gene therapy; vaccine.  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
XX W0200277021-A2.  
PN  
XX  
PD 03-OCT-2002.  
XX  
XX 27-MAR-2002; 2002WO-IB002163.  
PF  
XX 27-MAR-2001; 2001GB-00007658.  
PR  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Masignani V, Tettelin H, Fraser C;  
DR  
XX WPI; 2003-040579/03.  
DR P-PSDB; ABU02622.  
XX  
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection,  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
XX  
XX Claim 6; SEQ ID NO 4401; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the Streptococcus pneumoniae type 4 strain genomic sequence appearing as AB556454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a Streptococcus nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a Streptococcus pneumoniae bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to streptococcus bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2489 identified coding region from the genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [fr.wipo.int/pub/published\\_pct\\_sequences](http://fr.wipo.int/pub/published_pct_sequences). (Updated on 27-OCT-2003 to

CC	standardise OS field)
XX	Sequence 1515 BP; 433 A; 306 C; 351 G; 425 T; 0 U; 0 Other;
SQ	Query Match 19.1%; Score 287.4; DB 10; Length 1515; Best Local Similarity 51.7%; Pred. No. 1.5e-58; Matches 763; Conservative 0; Mismatches 696; Indels 18; Gaps 4;
QY	11 AACGCGTTTTGGAAATTGTCTCACCCCCCACCAGTTTTTCGGGTGCGTGGGGGATTCGGG 70
DB	8 AACGTCAAAGTGGTGTGATGCACATCTCTCTCTCCAGGAGCTTACGGAAATCGCAT 67
QY	71 CTCCTGGCGCGGAGGCCGCGGTTTTTGGACTGCTGGTCCGATGCGGAGCCGCTGGT 130
DB	68 CATTTGGTCAAAAGTGCTTACGACTTCGTGTGATTTCTTTGGTCCGTACAAAACAAGTTACT 127
QY	131 GGCAGGTCTTACCGCTGGGCCCCTACCAAGTTACGGCGACTCGCCGTACCAAGTCTTCTTCGG 190
DB	128 GGCAATCTTCCATTAGGAGCACTAGTTAGCGGGATTCTCTTACCAATCTTCTCTCAG 187
QY	191 CTTTTCGCGGTAAACCCGTATTGTTGTGACCCCGAGATGCTGATTGAAAAAGGCTGGCTGG 250
DB	188 CCTTCGCAGGAAACACTCATTTTATCGATTTAGATAATCTTTGGTGGAGCAAGTTTGTGG 247
QY	251' AACAAAGCGAAGGCCGCCCGCGGTATCCGACCCAGCGTGTGATTTATGGTGGCTTTTACC 310
DB	248 AAGCAAGTGACCTTGAAGGAGTTGACTTTTGGTAGGATGCTCTGAAGTTGACTATGCTA 307
QY	311 AGACCCGCTGGCCCTCTGTTGGCGGGGCTTTTCGGGGGTTTCGGGCAAGGGCTTCGCGCC 370
DB	308 AAATCTACTATGCAGCTCGTCTCTTTTGAAGAAAGCGGTGAACGTTTCTTTGAAGTCG 367
QY	371 AGGATAAAGACCCGACTGGAAGCTTTTATCGAGGCGGAGCGCTTCTGGCTGGAAGACTATG 430
DB	368 GAGATGTTAAAGATTTTGAGAAATTTGCTCAAGAACACCAATCATGCTGCTTGAAGCTCTTTG 427
QY	431 CGCTCTTTATGSCCTCAAGACCCCGTTTGTAGCGGCAAGCCCTCGGAACGAGTGG--AGCC 487
DB	428 CTGAGTATATGGCTATCAAGAAGATTTTGTGACAAATCTTTGCTTGAATGGTGAATGGAAG 487
QY	488 CCGAGCTGCGGACCGCTGAAACCGGCTGCTGGCGGAGGGCCCGTGAAGAGCTGCGCCGAGG 547
DB	488 CAGATGCTGCTGCTGTAAGCTTCAGGACATGGAAGATATCGTGAAGCAATTTGGCAGACA 547
QY	548 AGGTGCGCCCTTTACAGTGTGATTTAGTGGCTTTTATCTGGAATGGGGCCAGACCAAG 607
DB	548 AGTTGGTTTACCACCGTGTGACTCAATACTTCTTCTTCCAACAATGGTGAATTTGAAG 607
QY	608 CTTATGCCGAATCCAAAGGGATTCAGATATCGGGATATGCCCATCTTTGTGSGCTTCG 667
DB	608 CTTACGCTAACGACCAACCAATCGAAATCTGTTGGGAGATCCCAATCTACGTAGCGGAG 667
QY	668 ATTCTCTAGATCTCTGGGCAACCGCGAGTACTTCTACTCGAGGCCGATGCGCAACCCCA 727
DB	668 ATTCAGTGTATGTGGGCAATTCACATCTTTCAAAACAGATGTCAATGTGAAGGCTA 727
QY	728 CCGTGGTGGCGGGCGTTTCGGGGGACTACTTCTCGAAACCGGCCAGCTCTTGGGGCAATC 787
DB	728 CTTGTATGCGAGATGTCGCCACCGACAGATGATTTTCTGAATGCTGTTGGGGTAAATC 787
QY	788 CGCTCTATCGCTGGGATGTGATGAAAGGGCAACTTTTGGCTGGTGCATTCGCCGATAA 847
DB	788 CAATCTATGACTGGGAAGCAATGGACAAAGACGGCTACAAATGGTGGATTTGAACCTTGC 847
QY	848 GGCAGTGGCTCAAGCAGTGCACCTGGTGGCATCGACCATCTTCGCGGGGTTTGAAGCT 907
DB	848 GTGAAGCTTCAAAATCTACGATATCGTTGTGATCGACCATCTTCGCTGCTTGAATCTT 907
QY	908 ACTGGGAGGTTCGCTTTGCGCGGCCCAATGTCTGTGGAGGGGGCTTGGGTCAAAGCCCCAG 967
DB	908 ACTGGGAATCCCTGCTGTTCCGATACAGCAGCACCTGGTGAAGTGGTGAAGAGGTCCAG 967
QY	968 GGGAGAAGCTGTTTGTGCTGGGTGGGGGCCCAACTGACGATGCGCCCATCATTCATTCGCGAAG 1027

Db 968 GTTAAAGCTTTTTCAGCGGTTAAGGAAGAACTTGGTGAGCTAAACATCATCGCAGAAG 1027  
 Qy 1028 ACCTGGGGGTATACACCCCGAGGTGGAGGCTTTGGCGGATGGCTTCGGGTCCCGGCA 1087  
 Db 1028 ACCTTGGCTTATGACAGATGAAGTATCGAATTCGCTGAACGTACTGGCTTCCCGAAG 1087  
 Qy 1088 TGAAGATTTGCAAGTTGCTTTTCCGGTGAGGACAAAGCCCTTTTGGCCCCACAACTACC 1147  
 Db 1088 TGAAGATTTCTCAATTTGCTTCAACCCAGAGAGCAAGCAATTTGATAGCCCACT--- 1144  
 Qy 1148 CCGGCACGGCAATGTGTGTGTGTAGCGGGAACCCAGACACACACACCTCCGGAT 1207  
 Db 1145 TGGCACTGCTAACTCACTTATGTACAGGAACACACAGATAACAATACGGTCTTGTGT 1204  
 Qy 1208 GGTTCGACCGCGCGGAGCGGCGGCTTCATCGGGCCCTACCTGCGCCGCTATG 1267  
 Db 1205 GGTACCGTAATGAGATGATGATGCG-----ACTCGTGAATGATGCTCGTTACA 1255  
 Qy 1268 GCATCCGTTGTTGTGGAATACAGAGTTCGGGGCGCTTGTATGAGCTGGCTTCAAAA 1327  
 Db 1256 CGAACCGTAAAGATACGAACAGTGGT---ACAGCTATGCTTCGTACATATTTTCA 1312  
 Qy 1328 GCCCGCCCAAGCTGCTATTGTCCTTTGCGAGGAGTGTCTGGGGTGGGCCCGGAGGCC 1387  
 Db 1313 CAGTTAGCTTTATGCAATTCGAATTCGAAGATTTACTAGAAATGGATGAGCAGCTC 1372  
 Qy 1388 GCATGAATCTCCCGGACGGCTGGGGGACAACTGGGCGGTGGCGTACGCCGAAGCGGACC 1447  
 Db 1373 GTATGAATCTCCCATCTACCCCTGGTGAAACTGGTCTTGGCGTATGACTGAAGATCAAT 1432  
 Qy 1448 TCGAGCCCGTCTGCGCGGCGGACTGCGGGGCCCTGGC 1484  
 Db 1433 TGACACCAAGCTGTCGAGGAAGTTTGTGACTTGAC 1469

## RESULT 9

AAV52298  
 ID AAV52298 standard; DNA; 5910 BP.  
 XX AC AAV52298;  
 XX DT 23-OCT-1998 (first entry)  
 XX DE Streptococcus pneumoniae genome fragment SEQ ID NO:165.  
 XX KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 XX KW computer readable medium; vaccine; pharmaceutical composition; ds.  
 XX OS Streptococcus pneumoniae.  
 XX PN W09818931-A2.  
 XX PD 07-MAY-1998.  
 XX PF 30-OCT-1997; 97MO-US019588.  
 XX PR 31-OCT-1996; 96US-0029960P.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;  
 PI Dougherty BA;  
 XX WPI; 1998-272225/24.  
 XX PT Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.  
 XX Claim 1; Page 1063-1066; 1409pp; English.  
 XX The present invention describes a computer readable medium which has the

CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the S. pneumoniae  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to  
 CC 391, identifying members of the library which contain sequences that  
 CC hybridise to the target sequence and isolating the nucleic acid molecules  
 CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an  
 CC organism, amplifying nucleic acid molecules whose nucleotide sequence is  
 CC homologous to amplification primers derived from the fragment of the S.  
 CC pneumoniae genome to prime the amplification and isolating the amplified  
 CC sequences. The computer readable medium can be used in a computer-based  
 CC system for identifying fragments of the S. pneumoniae genome of  
 CC commercial importance, or expression modulating fragments of the S.  
 CC pneumoniae genome. Products from the present invention can be used in  
 CC diagnosis kits and assays, and pharmaceutical compositions and vaccines  
 CC for S. pneumoniae

XX  
 SQ Sequence 5910 BP; 1783 A; 1137 C; 1141 G; 1848 T; 0 U; 1 Other;

Query Match 19.1%; Score 287.4; DB 2; Length 5910;  
 Best Local Similarity 51.7%; Pred. No. 2.1e-58;  
 Matches 763; Conservative 0; Mismatches 696; Indels 18; Gaps 4;  
 Qy 11 AACGCGCTTTTGGAAATTTGCTCCACCCACAGTTTTCGGGTGCTGGGGGATTTGGG 70  
 Db 108 AACGTCAAAGTGGTGTGTGATGCATCTCTCTTCCAGGAGCTTACGGAATCGGAT 167  
 Qy 71 CTCTGGCGCGAGCGCGGCTTTTGGACTGGCTGGCGGATGCGGAGCCCGCTGCT 130  
 Db 168 CATTTGGTCAAGTGTTCAGACTTCGTTGATTTCTTTGGTCCGTACAAAACACAGTTACT 227  
 Qy 131 GGCAGGTCTTACCGCTGGGCGCTTACCAAGTTACGCGGCTGCGCGTACCAAGTCTTCTCGG 190  
 Db 228 GGCATAATCTTCCATTAGGAGCACTAGTTAAGGGGATTCCTTACCATCTTCTCAG 287  
 Qy 191 CTTTTCGCGTAACCCGTAATTTGGTTGACCCCGAGATGCTGATTTGAAAGAGCTGGCTGG 250  
 Db 288 CCTTCGCGAGAAACACTCAITTTATCGATTTAGATATCTTGGTGGAGCAAGCTTTGTGG 347  
 Qy 251 AACAAAGCAAGCGCCCGCCCGCTATCCGACCCAGCGCTGGATTTAGGCTGGCTTTACC 310  
 Db 348 AAGCAAGTGACCTTGAAGAGTGTGACTTTGGTAGCGATGCGTCTGAAAGTGTACTATGCTA 407  
 Qy 311 AGACCCCGCTGGCCCTGTTGCGCGGGCTTTTCGGCGCAAGGGCTTCGGGCC 370  
 Db 408 AATCTACTATGCACTGCTCTCTTTTGAAGAAAGCGGTGAACGTTTCTTTGAGTGG 467  
 Qy 371 AGGATAGACCCGACTGGAAGCCTTTATCGAGCGCGAGCGCTTCGTGGTGAAGCAATAG 430  
 Db 468 GAGATGTTAAAGATTTTGAGAAATTTGCTCAAGACCAACCAATCATGCTTGGCTCTTTG 527  
 Qy 431 CGCTCTTATGCGCCCTCAACACCGGTTTGAAGCGAGCCCTGGAAGAGTGG---AGCC 487  
 Db 528 CTGAGTATATGGCTATCAAAAGAGTATTTTGGCAATCTTGTGTGACTGGAATGGCCAGATG 587  
 Qy 488 CCGAGCTGCGGACCGTGAACCGGCTGCGCCAGGCGCCGCGAGAGCTGCGCCGAGG 547  
 Db 588 CAGATGCTGCTGCTGTAAGCTTCAGCACTTGAAGCTATCGTGAAGCAATTTGCGACACA 647  
 Qy 548 AGGTGGCCCTTTTACGAGTGGATTCAGTGGCTTTTATCTGGAAATGGGGCCAGACCAAGG 607  
 Db 648 AGTTGGTTTACCACCGGTGACTCAATCTCTTCTTCCAAACATGGTTGAAATTTGAAAG 707  
 Qy 608 CCTATGCCGATCAAGGGGATTCAGATTTACGGGATATGCCCATCTTTTGTGCGCTTCG 667  
 Db 708 CTTACGCTAACGACAACCAACATCGAATCGTTGGGACATGCCCAATCTACGTAGCGGAAG 767  
 Qy 668 ATTCTCAGATGTCTGGGCCAACCCCGCAGTACTTCTACTCGAGGCGGATGGCAACCCCA 727

Db	768	ATTCAAGTATATGTGGCAATCCACATCTCTTCAAAACAGATGTCAATGGTAAGCTA	827
Qy	728	CGGTGGTGGGGGTTCCCGGGGACTACTTCTCCGAAACCGGACAGCTCTGGGGCAATC	787
Db	828	CTTGATTCGAGAGATGCCACAGATGATTTCTGTAATGCTGAGCTTTGGGTAAATC	887
Qy	788	CGCTCTATCGCTGGATGTGATGAAAGGACAACTTTGCTGGTGTGATTTGCCCGCAATA	847
Db	888	CAATCTATGCTGGAGCAATGACAAAGACGGCTACAAATGGTGGATTTGAACGCTTGC	947
Qy	848	GGAGTGGCTCGACAGTGCACCTGTGCGATCGACCACTTCCGGGGTGTGAAGCCT	907
Db	948	GTGAAGAGCTTCAAAATCTACGATATCTGATATCGACCACTTCCGGGGTGTGAAGCCT	1007
Qy	908	ACTGGGAGTGTCCGTTTGGCGGCGCCCAATGCTGTGGAGGGCGCTGGGTCAAAAGCCCGAG	967
Db	1008	ACTGGGAAATCCCTGCTGTTCCGATACAGACAGCACTGGTGGTGAAGGTCAG	1067
Qy	968	GGGAGAAAGCTGTTGCTGGGTCGGGCGCCAACTGAGCGATGCGCCCATCATTTGCCGAAG	1027
Db	1068	GTACAAGCTTTTGCAGCGTTAAGGAAGAACTTGGTGAAGTAAACATCATCGCAGAAG	1127
Qy	1028	ACCTGGGGTATACACCCCGAGTGGAGGCTTTGCGGATGGCTTCCGGTTCGCCGCA	1087
Db	1128	ACCTTGGCTTCATGACAGATGAAGTGATCGAAATTCGCTGAACGCTGCTTCCCGGAA	1187
Qy	1088	TGAAGATTTTGCAGTGTCTTTTCCGCTGAGGACAAAGCGCTTTTGGCCCGACAACTACC	1147
Db	1188	TGAAGATTTTGCAGTGTCTTTTCCGCTGAGGACAAAGCGCTTTTGGCCCGACAACTACC	1244
Qy	1148	CCGCGCAGGCAATGTGGTGTGTACAGCGGAAACCCAGCAACGACGACCAACCTGGAT	1207
Db	1248	TGGCACCTGTCTAAGTATGTATGATACAGGAAACACGATTAACATACGCTTCTGGTT	1304
Qy	1208	GGTTCGCAACCGCGGAGCGGCGGCTTCTGCGGCGCTTCTGCGGCTTCTGCGCTATG	1267
Db	1308	GGTACCGTAATGAGATGATGATGCG-----ACTCGTGAATGATGCTGCTGTTACA	1355
Qy	1268	GCATCGCTGTTGTCGGAATACGAGGTGCGGGCGCTTTGATCGAGTGGCTTCAAAA	1327
Db	1356	CGAACCGTAAAGATACGAACAGTGT---ACAGCTATGCTTCTGATACAGTATTTTCAT	1412
Qy	1328	GCCGCGCAAGCTGGCTATTGCTGCTTTGAGGACGCTGCTGGGCTGGGCGCCGAGGCC	1387
Db	1413	CAGTTAGCTTTATGGCAATGCAACTATGCAAGATTTACTAGAAATGATGAGCGAGCTC	1472
Qy	1388	GCATGAATCTCCCGAGCGCTGGGGACACTGGGCTGGGCTGCTACGCGGAGCGGACC	1447
Db	1473	GTATGAATCTCCCATCTACCCCTTGGTGGAAACTGGTCTTGGCGTATGACTGAAGATCAAT	1532
Qy	1448	TCGAGCCGGTCTGCGCGGAGCTGCGGGCGCTTGGC 1494	
Db	1533	TGACACAGCTGTGAGGAAGTTTGTGCTTACTTGTAC 1569	

RESULT 10			
ABSS6454_19/c			
Continuation (20 of 22) of ABSS6454 from base 1900001 (Streptococcus pneumoniae type 4)			
WP Sequence split into 22 fragments			
WP	Fragment Name	Begin	End
WP	ABSS6454_00	1	110000
WP	ABSS6454_01	100001	210000
WP	ABSS6454_02	200001	310000
WP	ABSS6454_03	300001	410000
WP	ABSS6454_04	400001	510000
WP	ABSS6454_05	500001	610000
WP	ABSS6454_06	600001	710000
WP	ABSS6454_07	700001	810000
WP	ABSS6454_08	800001	910000
WP	ABSS6454_09	900001	1010000
WP	ABSS6454_10	1000001	1110000
WP	ABSS6454_11	1100001	1210000

WP	ABSS6454_12	1200001	1310000
WP	ABSS6454_13	1300001	1410000
WP	ABSS6454_14	1400001	1510000
WP	ABSS6454_15	1500001	1610000
WP	ABSS6454_16	1600001	1710000
WP	ABSS6454_17	1700001	1810000
WP	ABSS6454_18	1800001	1910000
WP	ABSS6454_19	1900001	2010000
WP	ABSS6454_20	2000001	2110000
WP	ABSS6454_21	2100001	2162598
Query Match 19.1%; Score 287.4; DB 10; Length 110000;			
Best Local Similarity 51.7%; Pred. No. 4.3e-58;			
Matches 763; Conservative 0; Mismatches 696; Indels 18; Gaps 4;			
Qy	11	AAGCGCTTTTGGAAATTTTGTCTCCAGCCACCAAGTTTTCGGGTCGCTGGGGATTTGGG	70
Db	13861	AACGTCAAAAGTGGTGTGATGACATCTTCTTCCAGAGCTTACGGAATCGAT	13802
Qy	71	CTCTGGGCGGAGGCGGAGCGGTTTTTGACCTGGCTGGGCGGATCGGGAGCCCGCTGGT	130
Db	13801	CATTGGTCAAGTGTCTTACGACTTCTGTTGATTTCTTGGTCCGTACAAAACAGATTACT	13742
Qy	131	GGCAGGTCTTTACCGCTGGGCGCTTACAGTTTACCGGACCTCCGCTACCACTTCTTCGG	190
Db	13741	GGCAATCTTCAATTTAGGAGCACTAGTTACGGGATTTCTCTTACCAATCTTCTCTCAG	13682
Qy	191	CTTTTCCGGTAACCCGTTATTTGGTTGACCCCGAGATGCTGATGAAAGAGCTGGCTGG	250
Db	13681	CTTTCGAGCAACACTCATTTTTATCGATTTAGATATCTTGGTGGAGCAAGGTTTGTGG	13622
Qy	251	AACAAAGCAAGCGCCCGCCGCTATCCGACCCGAGCGGTGATTTATGGCTTTTACC	310
Db	13621	AAGCAAGTGACCTTGAAGAGTTGACTTTGGTAGCGATGCTGTAAGTTGACTATGCTA	13562
Qy	311	AGACCGCTGGCGCCCTGTTTGGCGCGGCTTTTCGGCGGTTTCGGGCAAGGGCTTCGGCC	370
Db	13561	AAATCTACTATGACGCTGCTCTTTTAGAAAAAGCGGTGAAACGTTTCTTTGAAGTCG	13502
Qy	371	AGATTAAGACCCGACCTGGAAGCCCTTTTCGAGCGCGAGCGCTTCTGGCTCGAGACTATG	430
Db	13501	GAGATGTTAAGATTTTGAATAATTTGCTCAAGAACCAACCAATCATGGCTTGGCTCTTTG	13442
Qy	431	CGCTCTTTTGGCCCTCAAGACCCGCTTTCAGCGCAAGCCCTCGAACGAGTGG---AGCC	487
Db	13441	CTGATATATGGCTATCAAAAGATTTTTCGCAATCTTGGTGGACTGATGGCCAGATG	13382
Qy	488	CCGAGCTGCGGACCCGCTGAAACCGGCTGCTGCGCGGCGCCGCTGAGGAGCTGGCCGAGG	547
Db	13381	CAGATGCTGCTGCTGCTTAAAGCTTACGACTTGAAGCTATCTGAGCAATTTGCGAGACA	13322
Qy	548	AGGTGGCCCTTTACGAGTGATTTAGTGGCTTTTTCATCTGGATGGGCGGAGCCAGG	607
Db	13321	AGTTGGTTTACCAACCGCTGACTCAATCTTCTTCTTCCAAACCAATGGTTGAAATGAAG	13262
Qy	608	CCTATGCGGATCCAGGGGATTCAGATTTACGGGATATGCGGATATGCCATCTTTTGTGGCTTCG	667
Db	13261	CTTACGCTAAACGCAACCAATCGAATTCGTTGGGACATGCCAATCTACGAGCGGAG	13202
Qy	668	ATTCTCTCAGATGCTGGGCGCAACCCGAGTATTTTCTACCTCGAGCGCGATGGCAACCCCA	727
Db	13201	ATTCAAGTATATGTTGGCAATCCCACTCTTCTCAAAAACAGATGTCAATGGTAAGGCTA	13142
Qy	728	CGGTGGTGGCGGCTTCCGGGAGCTACTTCTCGAAACCGGCGAGCTCTGGGCAATC	787
Db	13141	CTTGATTCGCGAGGATGCCCAACAGATGATTTCTGTAACCTGGTGGCTTGGGGTAAATC	13082
Qy	788	GGCTCTATCGCTGGATGCTGTAAGAGGACCACTTTTGGCTGGTGTGATTTGCCCGCATAA	847
Db	13081	CAATCTATGCTGGGAAGCAATGGCAAGAGGCTACAAATGGTGGATTTGAAGCTTGC	13022
Qy	848	GGCAGTCTGCTCAAGCAGTGCACCTGCTGGCGATTCGACCACTTCCGGGGTTCGAAGCT	907









QY 388 GAAGCCTTTATCGAGCGGAGCGCTTCTGGCTTGAAGACTATGCGCTTTTATGGCCCTC 447  
DB 244 GAGAAATTTGCTCAAGACACCAATCATGGCTTGTGCTTGTGAGTATATGGCTATC 303  
QY 448 AAGACCCGGTTTGA CGGCAAGCCCTGGAACAGTGGAGC---CCGAGCTGCGGACCGT 504  
DB 304 AAGAGCAATTTGACAATTTGCTTGGACTGAAATGGCCAGATGCAGATGCTCGTGTCTCGT 363  
QY 505 GAACCGGCTGCGCTGGCCAGGCGCGCTGAGAGCTGGCCGAGGAGGTGGCCCTTTACGAG 564  
DB 364 AAGACTTCAGCACTTGAAGAGCTATCGTGAGCAATTTGGCAGACAGTGGTTTACCAACCGT 423  
QY 565 TGGATTCAAGTGGCTTTTATCTGGAATGGGGCCAGACCAAGGCTATGCGAATCAAG 624  
DB 424 GTGACTCAATCTCTTCTTCCAAATGTTGAAATGAAAGCTTACGTACGACACAC 483  
QY 625 GGGATTCAAGATTTACGCGGATATGCCCATCTTTTGGCCCTTCGATTCCTCAGATGCTGG 684  
DB 484 CACATCGAAATCGTTGGGACATGCCAATCTACGTAGCGGAAGATTCAAGTGATATGG 543  
QY 685 GCGAACCGGAGTACTTCTTACGTGAGCGCATGGCAACCCACGCGTGGTGGCGGGCTT 744  
DB 544 GCAATCCACATCTCTTCAAAACAGAGATGTAAGTGAAGCTACTTGTATCGCAGGATGC 603  
QY 745 CCGGGGACTACTTCTCGGAACCGGCCAGCTCTGGGCAATCGCTCTATCGTGGAT 804  
DB 604 CCACCAATGAGTCTTCTGTAACCTGGTCACTTGGGGTAAATCAATCTATGATCGGAA 663  
QY 805 GTGATGAAAGGAGCAACTTTGCTGCTGATTCGCCGCAATAGGCAAGTGCCTCAAGCAG 864  
DB 664 GCAATGCAACAGCGCTACAATGTTGATGAAAGCTTGGTGAAGCTTCAAAATC 723  
QY 865 TGCACCTGTGTGCGATCGACACCTTCCGGGGTTTGAAGCTTACGTGGAGGTTTCCGTTT 924  
DB 724 TAGCATATCTGCTATCGACCACTTCCGTTGGCTTCGAATCTTACTGGGAAATCCCTGCT 783  
QY 925 GCGCGGCCCAATCTGTGGAGGGCGCTGGGTCAAGCCCGGAGAGAGCTGTTTGTCT 984  
DB 784 GGTTCGATACAGAGCACCTGGTGTGAGTGAAGGTCCAGGCTACAGCTTTTGTCA 843  
QY 985 GCGGTGCGGCGCAACTGAGCGATGCGCCCATCAATTGCCGAAGACTGGGGGTGATCACC 1044  
DB 844 GCGTTTAAAGGAAGAACTTGTGTGAGCTAAACATCATCGCAGAGACCTTGGCTTCATGACA 903  
QY 1045 CCGAGGTGAGGCTTTGCGCGATGCTTCCGCTGCTCCCGGATGAAGATTTCGATGTT 1104  
DB 904 GATGAAGTGAATCGAATTTGCTGAACGTAATGCTGCTTCCAGGAATGAAGATTCTTCAATT 963  
QY 1105 GCTTTTCCGTGAGGACAGCGCTTTTTCGCCCACTACCTCCCGGACGCGCAATGTG 1164  
DB 964 GCTTCAACCCAGAGACGAAAGCAATGATAGCCCACT---TGGCACCCTGTAACTCA 1020  
QY 1165 GTGGTGTACAGCGGACCCACGACAGACACCACTGGGATGGTTCGACCGCGCG 1224  
DB 1021 GTTATGTACACAGACACACGATACATACGTTCTTGGTGTGATACCGTATGAT 1080  
QY 1225 GAGCGGAGGCGCTTCATGCGGGCTTACTCGGCGCTGATGCGCATCGGTTGTTGTCG 1284  
DB 1081 GATGATGCG-----ACTCGTGATGATCATGGCTTGTACACGAAACCGTAAAGAATAC 1131  
QY 1285 GAATACAGGTGCGGGCGCTTTGATCGAGCTGGCTTCAAAAGCCGCGCAGCTGGCT 1344  
DB 1132 GAAACAGTGGT---ACACGCTATGCTTCGPACAGTATTTTTCATCAGTACGTTTATGGCA 1188  
QY 1345 ATTGTGCTTTGAGGACGCTGCTGGGCTGGGCGCCCGGCGCATGAATTCCTCCCGGA 1404  
DB 1189 ATTGCACTATGCAAGATTTACTAGAAATGGATGAGCGACTGATGATGAATCTCCATCT 1248  
QY 1405 CGGCTGGGGACAACTGGGCGTGGCTAGCGGAAGGCGACCTCGAGCCCGCTCTGGCC 1464  
DB 1249 ACCCTTGGTGAAGAACTGGTCTTGGCGTATGACTGAAGATCAATTGACACAGCTGTCGAG 1308

QY 1465 GCGGAGCTGCGGCGCTTGGC 1484  
DB 1309 GAAGGTTGCTTGACTTGAC 1328  
RESULT 13  
ADQ15021  
ID ADQ15021 standard; DNA; 1731 BP.  
XX ADQ15021;  
XX 23-SEP-2004 (first entry)  
XX DNA encoding potato amylo maltase polypeptide.  
XX Amylo maltase; transgenic; foodstuff; food additive; modifier;  
KW saccharide; (alpha)-1,4-glucan; cooked rice; Japanese confectionery;  
KW snack; noodle; Chinese dumpling; shao-mai; fishery kneading;  
KW refrigerated; baby food; pet; animal feed; drink; food supplement;  
KW cyclic glucan; potato; gene; ds.  
XX Solanum tuberosum.  
OS  
XX  
FH Key Location/Qualifiers  
FT CDS 1. .1731  
FT /\*tag= a  
FT /product= "Potato amylo maltase polypeptide"  
XX JP2004187674-A.  
XX  
XX 08-JUL-2004.  
XX 12-SEP-2003; 2003JP-00322319.  
XX 26-NOV-2002; 2002JP-00342966.  
XX (EZAK ) EZAKI GLICO CO LTD.  
XX (NIDE ) NEC CORP.  
XX WPI; 2004-528719/51.  
XX P-PSDB; ADQ15022.  
XX New mutated amylo maltase polypeptide having increased enzyme activity  
XX and reduced hydrolyzing activity, useful for manufacturing foodstuffs  
XX such as Japanese confectionery, noodles, baby foods, and food additives.  
XX Example 9; SEQ ID NO 12; 51pp; Japanese.  
XX The invention relates to a novel amylo maltase polypeptide comprising the  
XX amino acid sequence of a wild-type amylo maltase with a substitution,  
XX addition or deletion at a position, where the amino acid residue  
XX interacts with acarbose of amino acids other than the wild-type amylo  
XX maltase amino acid sequence. The invention further comprises: a nucleic  
XX acid molecule containing a nucleic acid sequence encoding the novel amylo  
XX maltase polypeptide; a vector containing the nucleic acid sequence; a  
XX cell containing the nucleic acid sequence; biological tissue containing  
XX the nucleic acid sequence; a transgenic organism containing the nucleic  
XX acid sequence; a foodstuff, a food additive or a modifier of foodstuff  
XX containing the nucleic acid sequence; a computer readable recording  
XX medium which contains the information of the nucleic acid sequence  
XX encoding the amino acid sequence of the novel amylo maltase polypeptide;  
XX and a cyclic glucan obtained by reacting the novel amylo maltase  
XX polypeptide on saccharides which have a linear structure of (alpha)-1,4-  
XX glucan. The novel amylo maltase polypeptide is useful for manufacturing  
XX foodstuffs such as cooked rice, Japanese confectionery, snacks, bakeries,  
XX noodles, Chinese dumpling, shao-mai, fishery kneading goods,  
XX refrigerating foods, baby foods, pet foods, feed for animals, drinks, and  
XX food supplements, where the method involves adding the novel amylo  
XX maltase polypeptide to the foodstuff before or immediately after heat  
XX processing of the raw material, where the novel amylo maltase polypeptide  
XX generates a cyclic glucan from the starch of the foodstuff. The novel  
XX amylo maltase polypeptide is useful for manufacturing glucan which has a  
XX cyclic structure of (alpha)-1,4-glycoside linkage, foodstuff materials

CC and food additives, modifier of foodstuffs, a drink or eatable  
CC composition, infusion solution or composition for adhesion, where the  
CC method involves reacting the saccharide containing the linear structure  
CC of (alpha)-1,4-glycoside linkage or its derivatives with the novel amylo  
CC maltase polypeptide. This polynucleotide sequence represents the DNA  
CC encoding a potato amylo maltase polypeptide of the invention.  
XX  
SQ Sequence 1731 BP; 485 A; 344 C; 405 G; 497 T; 0 U; 0 Other;  
Query Match 12.1%; Score 181.4; DB 12; Length 1731;  
Best Local Similarity 48.3%; Pred. No. 3.3e-33;  
Matches 589; Conservative 0; Mismatches 701; Indels 37; Gaps 5;  
21 TGGAAATTTGCTCCACCCACACGATTTTCGGGTGCTGGGGGATTTGGGGCTCTGGGGCG 80  
Db TGGAAATTTGCTCCACCCACACGATTTTCGGGTGCTGGGGGATTTGGGGCTCTGGGGCG 299  
81 CGAGGCCGAGCGGTTTGGACTGGCTGGCGGATGCGGAGCCGCTGGTGCGCAGGTCTT 140  
Db TCAGGCTTTTAAAGTTCTTGATGGCTTCATCTTGCTGGTGGCTCCCTTTGGCAGGTCT 359  
141 ACCGCTGGGCCCTTAC-----CAGTTACGGGACTCGCCGTACCGAGTCTTCTC 188  
Db TCCACTTGTACCGCTCGAAGAGAGAGCAATGAAGATGGATCACCTATTACGACACAGGA 419  
189 GGCTTTTTCGGGTAAACCGGTATTTGGTTGACCCCGAGATGCTGATTTGAAAAGGCTGGCT 248  
Db TGCAAATTTGTGAAACACACTCTCTGATTTCTTTGAAAGAGCTTTGATGATGTTTACT 479  
249 GGAACAAAGCAAGCGCCCGCGTATCCGACCCAGCGCTGATTTATGCTGGCTTTA 308  
Db GAAGATGGAGAGGTTCCGAGGCACTACCTACAGATCGTGTCAATTACTCGACTATATC 539  
309 CCAGACCCGCTGGCCCTTGTTCGGCGGCTTTTCGGGGGTTTCGGGCAAGGGCTTCGGC 368  
Db TGAGTAAAGATCCTTTTAAATACCAAGGACGAAAGAGGCTTCTCTCCAGTGAAGGGA 599  
369 CCAGGATAAGACCGACTGGAAGCTTTATCGAGCGCAGGCGTT-----CTGCTGGAA 423  
Db ACTG--AAAGACCAAGCTCGAGAATCTTCGCGGGATCCAAATATTTCCAGTTGGCTGGAG 657  
424 GACTATGCGCTTTATGCGCCCTCAAGACCCGTTTGAACGCGCAAGCCCTGGAACGAGTGG 483  
Db GATGCTGCTTATTTGCTGCCATAGACAACTCTGTAAACACTATTAGCTGTATGATTTGG 717  
484 AGCCCGAGCTGCGCGACCGTGAACCGGCTGCCCTGCGCAGGCGCGGTGAGGAGCTGGCC 543  
Db CTTGACCAATTGAAAATCGCAATCTTCGAGCTTAGAAGAGTTTATCAAGTGAAGAAAG 777  
544 GAGGAGTGGCGCTTTAGAGTGGATTCAGTGGCTTTTATCTGGAATGGGCGCAGACC 603  
Db GATTTATTGACATATTCAITGCAACACAGTTCTTTGTTCCAAACGACATATGAAAAGTT 837  
604 AAGGCTATGCGGATCCAAAGGGATTCAGATTTACGGCGATATGCCATCTTTGTGGCC 663  
Db CGTGACTATGACGATCCAAAGGAATCAGTATAATGGAGACATGCCAATATATTTGGA 897  
664 TTCGATTCCTCAGATGCTCGGCGCAACCCCGCAGTACTTCTACCTCGAGCGCGATGGCAAC 723  
Db TATCACAGTGTGATTTTGGGCCNACAAGAAACAAATTTTGTCTGAAATAGGAAGTTTC 957  
724 CCCACGTTGGTGGCGGGCTTCGCGGAGTACTTCTCCGAAACCGGCGCAGCTCTGGGGC 783  
Db CCTCTTATAGTATGTTGGTGTCTCTCCAGACGCTTTAGTGAACCTTGTGCACTATGGGGC 1017  
784 AATCGCTCTATCGTGGGATGATGAAAGGACAACTTTGCTGGTGCATTTGCCCGC 843  
Db AGCCCTCTCTATGATGGAAGGCCATGGAGAGGATGGAATTTTCATGTTGGGTAGCCGA 1077  
844 ATAGGCGAGTCTCAACGAGTGGCCACCTGGTGGCGCATCGACCACTTCCGCGGTTTGA 903  
1078 ATTCAACGTCACCGATCTTTTGTATGATTTAGATAGATCATCTTTAGAGGATTTGCT 1137

QY 904 GCCTACTGGAGGTTCCGTTTGGCGGCCCAATGCTGTGTGGAGGGCGCTGGTCAAAAGCC 963  
Db 1138 GGATTTTGGCTGTTCTTCTGAGGAAAATTTGCAATTTCTGGACGGTGGAGGTGGGA 1197  
QY 964 CCAAGGGAGAGAGCTGTTTGTGCGGTGCGGGCCCAACTAGCGATGCGCCCATCATTTGCC 1023  
Db 1198 CTTGGAAGAGCCTTTGTTTGTATCTTACAAAGCTGTTTGGGAAGATCAATATATATAGCA 1257  
QY 1024 GAAGACTGGGGTGATCAACCCCGAGGTGGAGGCTTTGCGCGATGGCTTCGGGTTCCCC 1083  
Db 1258 GAAGACTTTGGGAGTAAATACCGAGGACGTTTGTTCAGCTAAGAAAGTCCATTGAGGCACT 1317  
QY 1084 GGATCAAGATTTTGCAGTTTGTCTTTTCCGTGAGGACAAACGCTTTTTCGCCCCCAAC 1143  
Db 1318 GGAATGGCTGACTCAGTTTGCATTTGGCAGTGAAGCTGAACCTCATTTACCTCAC 1377  
QY 1144 TACCCCGCGACGCGCAATGTGTGTACAGCGGAAACCAACGACACGACACCCCTG 1203  
Db 1378 AATCATGAGCA---GAACCAAGTAGTATATCTGGAACACATGACAAATGATACGATCCGA 1434  
QY 1204 GATGTTCCGACCCCGCGGAGGCGGCGGCTTCATGCGGGCTACCTGGCGCGC 1263  
Db 1435 GGTGTTGGGATCTTTGCCACAGGAAGAGAAATCCAAATGATCTAAAGTATTTATC--- 1490  
QY 1264 TATGGCATCGGTTGTTTGTTCGGAATACGAGGTGCGCGGCTTTTGTATCGAGCTGGCCTTC 1323  
Db 1491 -----AATATTGAGGAAGAGAAATATCACGGGGCTTGTATCGAAGGTGCAGTT 1539  
QY 1324 AAAAGCCCGCCCAAGCTGCTATTGTGCTTTTCAGGACCTGTCTGGGGCTGGGCCCCGAG 1383  
Db 1540 TCTTCTGTAGCCCGTATGCAATATATACCGATGCAAGATGTTCTTGGGCTTGGAGTGAT 1599  
QY 1384 GCCGATGAACTTCCCGGACGCTGGGGGACAACTGGGCGTGGCG 1430  
Db 1600 TCCAGATGAACATTCAGCACTCAGTTTGGAAACTGGAGTTGGAG 1646  
RESULT 14  
ABN69262  
ID ABN69262 standard; DNA; 1491 BP.  
XX AC AC  
XX AC  
XX AC  
DT 01-JUL-2002 (first entry)  
XX  
DE Streptococcus polynucleotide SEQ ID NO 6437.  
XX  
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
XX Streptococcus pyogenes.  
XX  
XX WO200234771-A2.  
XX  
XX 02-MAY-2002.  
XX  
XX 29-OCT-2001; 2001WO-GB004789.  
XX  
XX 27-OCT-2000; 2000GB-00026333.  
XX  
XX 24-NOV-2000; 2000GB-00028727.  
XX  
XX 07-MAR-2001; 2001GB-00005640.  
XX  
XX (CHIR-) CHIRON SPA.  
XX (GENO-) INST GENOMIC RES.  
XX Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;  
XX Tettelin H;  
XX WPI; 2002-352536/38.  
XX DR P-PSDB; ABP28631.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection or

PT disease caused by Streptococcus bacteria, such as meningitis, and for  
 PT detecting a compound that binds to the protein.

XX Claim 7; Page 3805; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/SBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins

XX SQ Sequence 1491 BP; 474 A; 247 C; 309 G; 461 T; 0 U; 0 Other;

Query Match 11.5%; Score 173; DB 6; Length 1491;

Best Local Similarity 47.1%; Pred. No. 3.2e-31;

Matches 56; Conservative 0; Mismatches 635; Indels 3; Gaps 1;

QY 9 CCAACGGCGCTTTTGGAAATTTTGTCTCCACCCACACAGTTTTCGGGGTCTGGGGGATTGG 68  
 DB 6 CAAAGGTGAAGCGGTATCTTAATGATATTAAGTTCCCTGCTGGGAATTTGGGATAGG 65  
 QY 69 GGCTCTGGCGCGAGCGCGGTTTGGACTGGTGGCCGATGGGAGCGCCGCTG 128  
 DB 66 CACTTTTGGAAATTCAGCTTTTGAATTTGTGATTTTCTAGCAGAGACGAACAACTA 125  
 QY 129 GTGCGAGTCTTACCGTGGGCCCTTACAGTTAGCGGACTCGCGGTACCAGTCTCTC 188  
 DB 126 TTGGCAATTTCTGCTTTAACACGACGAGTTTGGAGATTTCTCTTATCAGTCAATTTTC 185  
 QY 189 GGCTTTTCCGGTAACCCGCTATTTGGTTGACCCCGAGATGCTGATGAAAGCTGGCT 248  
 DB 186 AGCTATTGCTGGGAACACACATTTTCAATTTTGAATTTGCTAGTATGATGAATTTT 245  
 QY 249 GGAACAAAGGAGCGCCCGCGGTATCCGACCCAGCGGTGATGCTGGCTTTTA 308  
 DB 246 AGAAGCAGCAGATTTGTGTGATATTAATTCATTCGGCACAAATCTTGAAGCAGTAGACTATGC 305  
 QY 309 CCAGACCCGCTGGCCCTGTTGCGGGGCTTTCGGGGTTCGGCAAGGGCTTCGCG 368  
 DB 306 TCAGCTTTTCAAGTTAGACGTCACTTTTGAATAAGCAGTAGAGACTTTTGTGCTGA 365  
 QY 369 CCAGGATAAGACCCGACTGGAAGCTTTATCGAGCCGCGAGCGCTTCTGGCTGGAAGACTA 428  
 DB 366 ACAAGAAATGTGTAAATTAGAGCTTTTGAACAGCTTCTAGCTGGTTAACTGATTT 425  
 QY 429 TGGCTCTTTATGSCCCTCAAGACCCCGTTTGAACCGCAAGCCCTTGAACGAGTGGAGCCC 488  
 DB 426 TGCTGAATTTATGCTTTTAAAGAAATATTTTAAATAAAGCCCTTACAAGATTGGGACGA 485  
 QY 489 CGAGTGGCGC--GACCGTGAACCGCTGCGCCCGCGGCGCTGAGGAGCTGGCCGA 545  
 DB 486 CGAACTGTCTATTAAACGCCAAGAAGACTCACTTAATAATTACCGTAGTGTCTTGCAA 545  
 QY 546 GGAGTGGCCCTTTACAGTGGATTCAGTGGCTTTTATCTGGAATGGGCGCCAGACAA 605  
 DB 546 AAAGATTACCTATCATTAAGCTGTCTCACTACTTCTTTTCAACAAATGGTCAGCTTGAA 605  
 QY 606 GGCTATGCGCAATCCAGGGGATTCAGATATTCGGCGATATGCCCATCTTTTGGCGCTT 665  
 DB 606 AACTTAGCTTAACCATAAAGGGATAGAAATTTATGGTGAATATGCTTATTTATGTTCTGC 665

QY 666 CGATTTCTCAGATGCTGGGGCAACCCGACGACTTCTTACCTCGAGGCCGATGCAACCC 725  
 DB 666 GGACAGTGTAGAGTTTGGACGATCCCTGAGCTATTTAAAGTAGATAGTATAAAAAGCC 725  
 QY 726 CACGGTGGTGGCGGGCTTCGCGGACTACTTCTCCGAAACCCGCCAGCTCTTGGGGCAA 785  
 DB 726 ATTGTTTATTGACAGTGTACCAGCAGATGGCTTTAGTGAAGATGGTCACTGTGGGAAA 785  
 QY 786 TCCGCTCTATCGCTGGGATGTGAAAGGACAACTTTCCTGCTGGTGGTGCATTTGCCCGAT 845  
 DB 786 TCCAACTACAAATTTGGTCTGCTCATGAAAGTCTAAATTTTGCATGGTGAATTTATCGTAT 845  
 QY 846 AAGCAGTCTCAAGCAGTGCACCTGGTGGCATCGACCACCTTCCGCGGTTTGAAGC 905  
 DB 846 TCAAGAAAGTTTCAAAT 905  
 QY 906 CTACTGGAGGTTTCCGTTTGGCGGCGCCAAATGTGTGGAGGGCGCTGGGTCAAGCCCC 965  
 DB 906 TTTTGGGAAATTCCTGCTGGTGAATAAACTGCTAGAAATGGCCATTGGGCTCAGCCCC 965  
 QY 966 AGGGAGAACTGTTTCTGCTGGTGGGCGCCAACTAGCGATGCGGCCATCATTTGCCGA 1025  
 DB 966 AGGTATAGCAATTTTCTGCGGTTCTGAGGCGCTTAGGAGAGCTCCCTATCATCGCAGA 1025  
 QY 1026 AGACCTGGGGTGTATCACCCCGAGGTGGAGGCTTTTCGGGATGCTTTCGGGTTTCCCGG 1085  
 DB 1026 AAACCTGGGTTATATGATGAAAGCAGACAGCTATTAGCATCAACAGGCTTTCTCTGG 1085  
 QY 1086 CATGAAGATTTTGCAGTTCCTTTTCCGGTGGAGCAACAGCCCTTTTTCGCCCAACAATA 1145  
 DB 1086 CATGAAGATTTTGAATTTGGTTTATTTGATATAACGAGTCAAAAGTATAGATTTTACCACA 1145  
 QY 1146 CCCGGCAGCGCAATGTGTGTACAGGGAAACCCAGCAACAGCACACCCCTGGG 1205  
 DB 1146 TTTACTATGACCGTAATTTGTGCTGTACACAGGAGCTCATGATAACGAGGTGTTAACGG 1205  
 QY 1206 ATGCT 1210  
 DB 1206 ATGCT 1210

# RESULT 15

ABN69261

ID ABN69261 standard; DNA; 1494 BP.

XX AC ABN69261;

XX DT 01-JUL-2002 (first entry)

XX DE Streptococcus polynucleotide SEQ ID NO 6435.

XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
 XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.

XX OS Streptococcus agalactiae.

XX PN W0200234771-A2.

XX PD 02-MAY-2002.

XX XX 29-OCT-2001; 2001WO-GB0004789.

XX XX 27-OCT-2000; 2000GB-00026333.

XX XX 24-NOV-2000; 2000GB-00028727.

XX XX 07-MAR-2001; 2001GB-00005640.

XX XX (CHIR-) CHIRON SPA.

XX XX (GENO-) INST GENOMIC RES.

XX PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;

XX PI Tettelin H;

DR WPI; 2002-352536/38.  
 DR P-PSDB; ABP28630.  
 XX New Streptococcus protein for the treatment or prevention of infection or  
 PT disease caused by Streptococcus bacteria, such as meningitis, and for  
 PT detecting a compound that binds to the protein.  
 XX  
 XX Claim 7; Page 3805; 4525pp; English.  
 PS  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus (Streptococcus agalactiae) or group A streptococcus (Streptococcus  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and anti-inflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins  
 XX  
 XX Sequence 1494 BP; 494 A; 238 C; 316 G; 446 T; 0 U; 0 Other;  
 Query Match 11.3%; Score 169.4; DB 6; Length 1494;  
 Best Local Similarity 48.9%; Pred. No. 2.3e-30;  
 Matches 543; Conservative 0; Mismatches 561; Indels 7; Gaps 3;  
 QY 11 AACGCGCTTTGGAAATTTGCTCCACCCACAGTTTTCGGGTCGCTGGGGATTGGGG 70  
 DB 8 AACGTGCAAGTGGTGTCTTAATGCAATCATCTCTCTACAGGTGATTTAGTATGGAA 67  
 QY 71 CTCTCGGCGCGAGCGCGAGCGGTTTGGACTGCTGCGCGATCGCGAGCGCCGCTGGT 130  
 DB 68 CATTGGAGCAGAGCGCTACGCTTTTGTGATTTTGTGCGAACTGATCAAAAATTTT 127  
 QY 131 GCGAGGCTTTACCGTGGGCCCTACAGTTAGCGGACTCGCGGTACAGTCTTCTCGG 190  
 DB 128 GGCATAATTTCTCCCTTAACAACAAGCTTCGGGACTCTCTTACCAGTCTTTTCTG 187  
 QY 191 CTTTTCGGGTAAACCGTATTTGGTTGACCCGAGATCTGATTGAAAAGGCTGGCTGG 250  
 DB 188 CAGTAGCTGGTAATACACATTTAAATGATTTTGTGTTGTAACGCTTGAGGTTTCATTT 247  
 QY 251 AACAAAGCGAAGCGCCCGCGGTATCCGACCCAGCGCGTGGATTATGGCTGGCTTTACC 310  
 DB 248 CGAAGATGA--CTACCAAAATATTAGTTTGGACAGGATCCAGAAGTTGTTGATTATGC 305  
 QY 311 AGACCGGCTGGCCCTGTTTGGCGGGGCTTTTCGGGGGTTTCGGGCAAGGCTTCGGCCC 370  
 DB 306 TGGTCTGTTTGAATAAACCGCGCTTGTAGAAAAGCAGTTTAAAAATTTCTTCAAGA 365  
 QY 371 AGATAAG--ACCGACTGGAAGCTTTATCGAGCGCGAGCGCTTCTGGCTGGAAGACTA 428  
 DB 366 AGAGAGAGCTACGAGAATCTATCTGATTTCTTGAAGAAGAAAATTTGGGTAACTGATTT 425  
 QY 429 TCGGCTCTTTATGGCCCTCAAGACCCGGTTTGAACGCAAGCCCTGGAAACGAGTGGAGCCC 488  
 DB 426 TGTCTGATTTATGGCGATCAAGAACATTTTGGTATAGGCGCTTCAAGAATGGGATGA 485  
 QY 489 CGAG--CTGCGCGACCGGTGAACCGGCTGCCTTCGGCCAGCGGCCGAGAGCTGGCCGA 545  
 DB 486 CAAGGCTATTATACGCGCGGAGAGAAGCCCTTAGCAGGATATCGTCAAAAGCTTAGTGA 545  
 QY 546 GGAGGTGGCCCTTTACAGTGGATTACAGTGGCTTTTATCTGGAATGGGCGCAGACCAA 605  
 DB 546 AGTGATAAAATATCATGAAGTAACGCAATATTTCTTTTACAAACAATTTGTTGATTA 605

Search completed: January 14, 2006, 11:07:47  
 Job time : 888 secs

QY 606 GGCCTATGCCGAATCCAAGGGGATTTCAGATTATTCGGCGATATGCCCATCTTTGTGGCCTT 665  
 DB 606 AGAATATGCTAAATGATTAAGGGATTCAATATTCGGTGATATGCCAATCTACGTTCTGC 665  
 QY 666 CGATTCTCTCAGATGTCTGGGCCCAACCGCGAGTACTTCTACCTCGAGGCCGATGGCAACCC 725  
 DB 666 CGATAGTGTAGAAGTTTGGACAATGCTGAACCTGTTTAAATATAGATCGCGACAACAACC 725  
 QY 726 CACGGTGGTCCGGCGGCTTCGGCGGAGTACTTCTCGAAACCGCCAGCTCTCGGGCAA 785  
 DB 726 ACTTGCTATTGCTGGTGATACCGCGGATGATTTTCTGACGATGTCAGCTTTGGGGCA 785  
 QY 786 TCCGCTCTATCGCTGGGATGTGATGAAAGGACAACTTTGCTCTGGTGGTCAATTCGCCGAT 845  
 DB 786 TCTATTACAAATTTGGGATTTACCAAAAGAGTCAGATTTTGACTGGTGGATTTATCGTAT 845  
 QY 846 AAGGCAGTCCGCTCAAGCAGTGGCCACTGCTGGCGCATCGACCACTTCCGCGGTTTGAAGC 905  
 DB 846 TCAATCTGTTGTTAAGATGTATGACTATTTTGGCTATTGATCATTTTAAAGGTTTTCAGA 905  
 QY 906 CTACTGGGAGGTTCCGTTTGGCGGCCCAATGCTGTGGAGGGCGCTGGGTCAAAGCCCC 965  
 DB 906 TTATTGGGAAATTCGAGGAGATTTATCAACAGCTTAATGATGGCTCATGGCAACAGCTCC 965  
 QY 966 AGGGGAGAAGCTGTTTCTGCTGGTGCGGGCCCACTGAGCGATCGCGCCATCATTTGCCGA 1025  
 DB 966 CGGTCCAGAAATTTATTCGAATATCAAGAAATAATTAGTGTATTTACCTATTATTGCTGA 1025  
 QY 1026 AGACTGGGGGTGATCACCCCGAGGTGGAGGCTTTGGCGGATGGCTTCGGGTTCCCGCG 1085  
 DB 1026 GAATTTAGGCTATATTGATGAGAGAGCAGAGATTAATGCTGGAACAGGGTTCCCGAG 1085  
 QY 1086 CATGAAGATTTTGCAGTTTGTCTTTTCCGGT 1116  
 DB 1086 TATGAATAATCATGGAATTCGGTTTTTATGAT 1116

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2006, 14:04:33 ; Search time 39 Seconds  
(without alignments)  
1236.015 Million cell updates/sec

Title: US-10-618-976-2

Perfect score: 2714

Sequence: 1 MQLQAFGILLHTPTFFGRW.....DLFGLAAGLALAEASQRA 501

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_80:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1305.5	48.1	502	2 AH2289	4-alpha-glucanotra
2	1222.5	45.0	505	2 S74548	4-alpha-glucanotra
3	1194	44.0	505	2 E95246	4-alpha-glucanotra
4	1194	44.0	505	2 B98111	4-alpha-glucanotra
5	1100	40.5	485	2 E70363	4-alpha-glucanotra
6	1035.5	38.2	576	1 A45049	4-alpha-glucanotra
7	727	26.8	506	2 F70120	4-alpha-glucanotra
8	451	16.6	526	2 G72091	4-alpha-glucanotra
9	451	16.6	526	2 F86531	4-alpha-glucanotra
10	430.5	15.9	527	2 G71557	glucanotransferase
11	400.5	14.8	527	2 A81712	probable glucanotr
12	341.5	12.6	698	2 AC0016	4-alpha-glucanotra
13	320.5	11.8	724	2 G70328	4-alpha-glucanotra
14	313.5	11.6	684	2 F83028	probable 4-ALPHA-G
15	302.5	11.1	693	2 A10396	hypothetical prote
16	302	11.1	694	1 C65137	4-alpha-glucanotra
17	300.5	11.1	738	2 T00748	4-alpha-glucanotra
18	293	10.8	694	2 A86007	4-alpha-glucanotra
19	293	10.8	694	2 B91161	4-alpha-glucanotra
20	292.5	10.8	726	2 B82511	4-alpha-glucanotra
21	271.5	10.0	699	2 H64118	4-alpha-glucanotra
22	249	9.2	489	2 H86711	4-alpha-glucanotra
23	122	4.5	1271	2 E83350	hypothetical prote
24	116	4.3	554	2 H95922	hypothetical membr
25	114	4.2	585	2 T36060	probable oxidoredu
26	111	4.1	595	2 H95006	beta-galactosidase
27	110	4.1	595	2 C97879	beta-galactosidase
28	105.5	3.9	618	2 D87651	prolyl oligopeptid
29	104.5	3.9	629	2 E82624	thiamin biosynthes

## ALIGNMENTS

### RESULT 1

AH2289

4-alpha-glucanotransferase [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004

C;Accession: AH2289

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, D.N.A. Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium A

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AH2289

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-502 <KUR>

A;Cross-references: UNIPROT:Q8YQ90; UNIPARC:UPI000000CE92F; GB:BA000019; PIDN:BA075570.;

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr3871

C;Superfamily: 4-alpha-glucanotransferase

Query Match 48.1%; Score 1305.5; DB 2; Length 502;

Best Local Similarity 49.9%; Pred. No. 5.1e-96;

Matches 252; Conservative 80; Mismatches 162; Indels 11; Gaps 5;

QY 1 MQLQAFGILLHTPTSPFGWIGALGREAFRLDADAGARWQVLPGLPTSYGDSYQ 60

DB 1 MPFIRSSGVLLHTPTSCPSRFGIGDLGLEAYKIDFLKSYQQWQVLPGLPTGYGNSPYM 60

QY 61 SFSAFAGNLYVDPEMLIEKGMLEQSEAP--PPYPTQVDYQWLYQTRWPLLRFAFAGR 118

DB 61 SYSALAGNHLISPEKLLDEGLLSDEDFAHLENFNEKVDQVAPIKIQLKKACENFK 120

QY 119 ABAQAQDKTRLEAFTEAFERFMLEDFALFALXTRDQKFWNEWSPRLRREPAAALARE 178

DB 121 TKASPLQKGFAGFCETKSYWLDYALFALXKDTKSSSWHTWEPALAKREPDALEKQVR 180

QY 179 ELAEEVALYEWIQWLFYLEWGTQKVAESKGIQITGDMPIFVAFDSSDYVWNPQYFVL-E 237

DB 181 QUTDEIFYKFIQYEFFRQWSELKSYANNRGLEIIGDIFYVHDSADVWNPDIICLDE 240

QY 238 ADGNPTVAGVPRDYFSETGQLWGNPLYRWDMYMRDNFAWCIARIQSLKQCHLVRIHF 297

DB 241 ETGEVALMAGVPPDYFVSATGQLWGNPVYVWEELOKQDFKWWQVRFEAMDLYVDVIRIDHF 300

QY 298 RGFEAYWVPGFRPNVAGRWKAGEKLFANVRAQLSDAPIAEDLGVITTEVEALROG 357

DB 301 RGFEAWTVPOGEETANNGEWVTAPEELFDAIKQKGLKPLVLAEDLGVITTEVEALROK 360

QY 358 GFPGMKILQAF--SGEDNAPLPHNYPAGHNVVYSGTHDNTTGLGWFTAPAEAFNR 416

DB 358 GFPGMKILQAF--SGEDNAPLPHNYPAGHNVVYSGTHDNTTGLGWFTAPAEAFNR 416

QY 358 GFPGMKILQAF--SGEDNAPLPHNYPAGHNVVYSGTHDNTTGLGWFTAPAEAFNR 416

DB 358 GFPGMKILQAF--SGEDNAPLPHNYPAGHNVVYSGTHDNTTGLGWFTAPAEAFNR 416



Db 361 YEPGKVLQFAFGSDPGNPFLLPFNYSR--NFVVVTGTHDNDTTVGWENTANDYENKNLL 418

QY 417 AYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFFPGLGDNWAW 476

Db 419 LYLQ-----CISPEGIHWDLIRLASSVANQAIIVPLQDILGLGNEARWNFFSIAGSNWAW 473

QY 477 RYAEGLDFGLAAGLALAEASQRA 501

Db 474 RYDFAVLTDELSDRLKILTKLYGRA 498

RESULT 2

S74648

4-alpha-glucanotransferase malQ - Synechocystis sp. (strain PCC 6803)

N:Alternate names: hypothetical protein sll1676

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004

C:Accession: S74648

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis s.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S74648

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-505 <KAN>

A:Cross-references: UNIPROT:P72785; UNIPARC:UPI000012EBSE; EMBL:D90900; GB:AB0001339; NID

A>Note: the nucleotide sequence was submitted to the EMBL data library, June 1996

C:Genetics:

A:Gene: malQ

C:Superfamily: 4-alpha-glucanotransferase

Query Match 45.0%; Score 1222.5; DB 2; Length 505;

Best Local Similarity 48.6%; Pred. No. 2.1e-89;

Matches 242; Conservative 76; Mismatches 169; Indels 11; Gaps 6;

QY 4 QRAFGLLHPTSPFGRWGIGALGRAERFLDLADAGARWQVLPGLPPTSYGDSYQSF 63

Db 4 KRCSGILLHPTSLPRFGLDGDGAFQIDFLADQSVQWILLPLGPTGFNSYLCYS 63

QY 64 AFAGNPYLVDPEMLIEKWLQES--EAPPYPTQVDYGLWYQTRWPLLRRAFAFRARA 121

Db 64 ALAINPWLISLRLAEGLPPLSDQAPPFTNPRVDYQAIYKSVQLKQAFQRTNI 123

QY 122 SAQDKTRLEAFTEARFWEIDYALFMALKTRFDGKPNWSPDLDRREPALARAREELA 181

Db 124 ELAIEQFAEFQCAQSDMLADYALFMAIKEAHNGAGHWDKDIARWEPEALKIWDRLK 183

QY 182 ESVALYEWIOWLFYLEWQGTAKYAESKGIQIGDMPIFVAFDSSDVWANPQVLYLEAD-G 240

Db 184 TEVLHQFQFQIFGFRQWQEKYANQRHIAIFGDLPIYVAHSDAWANPENFLDPTG 243

QY 241 NPTVAGVPRDYFSETGQWGNPLRYWDMVNERDNFANCIARIQSLKQCHLVRIIDHFRG 300

Db 244 EAMMAGVPPDFVSATGQWGNPVYDWTLKATGFPAWKRFKANKLYLDIVRIDHFRGF 303

QY 301 EAYWEVFPGRPNAVEGRWVKAPGEKLPAAVRAQLSD-APIIAEDLGVTTPVEALRDGFG 359

Db 304 ESYWGVQCEKTAENGEMYPAPGKEPFQALGKALGDNLPDVAEDLGVTTPVEALRDDEFN 363

QY 360 FPGMKILOFAP-SEGDNALPHNYPAGHNVVYVSGTHDNDTTLGFRTAPEAERAFMAY 418

Db 364 FPGMKVLHFAFSDRGNPPLPFNY-SNGNAVVTGTHDNDTTVGWFOERSEDDQQKVINY 422

QY 419 LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFFPGLGDNWAW 478

Db 423 LG-----CVCNEGIHWSLIRLASSVAALAIPLQDILGLGSDCRNLFPGTAAGNWGRY 477

QY 479 AEGDLEFGLAAGLALAE 496

Db 478 HPDQLNDWLSCHLSFITE 495

RESULT 3

E95246

4-alpha-glucanotransferase [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004

C:Accession: E95246

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; He: on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle: neon, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Iofus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison

A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: E95246

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-505 <KUR>

A:Cross-references: UNIPROT:P29851; UNIPARC:UPI0000051B0A; GB:AB005672; PIDN:AAK76166..

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP2107

C:Superfamily: 4-alpha-glucanotransferase

Query Match 44.0%; Score 1194; DB 2; Length 505;

Best Local Similarity 48.1%; Pred. No. 3.8e-87;

Matches 242; Conservative 77; Mismatches 168; Indels 16; Gaps 9;

QY 4 QRAFGLLHPTSPFGRWGIGALGRAERFLDLADAGARWQVLPGLPPTSYGDSYQSF 63

Db 3 KRCSGVLHHSLLPFGAYGIGSGFQSDYDFVFLVTKQRYWQIILPLGATSYGDSYQSF 62

QY 64 AFAGNPYLVDPEMLIEKWLQESAPP---PYPTQVDYGLWYQTRWPLLRRAFAFRAR 120

Db 63 AFAGNTHFIDLDILVEQGLLEASDLGVDFGSDASEVDYAKIYYARRPLLEKAVKRFEEV 122

QY 121 ASADQKTRLEAFTEARFWEIDYALFMALKTRFDGKPNW-SPELRDREPALARAREE 179

Db 123 GDVKD---FEKFAQDNQSWLELFAFYNAIKFYFDNLAWTEWPDADARAKASALESYREQ 179

QY 180 LAEEVALYEWIOWLFYLEWQGTAKYAESKGIQIGDMPIFVAFDSSDVWANPQVLYLEAD 239

Db 180 LADKLVRHVTVQYFFQWQKLKAYANDNHIEIVGDMPIYVAEDSSDMWNPHEFKTDVN 239

QY 240 GNPTVAGVPRDYFSETGQWGNPLRYWDMVNERDNFANCIARIQSLKQCHLVRIIDHFRG 299

Db 240 GKATCIAGCPDEFVITGQWGNPIYDWEAMDKGYKMWIERLRESFKIYDIDRIDHFRG 299

QY 300 FEAYWEVFPGRPNAVEGRWVKAPGEKLPAAVRAQLSDAPIIAEDLGVTTPVEALRDGFG 359

Db 300 FESYWEIPAGSDTAAPGEWVKPGYKLPAAVKEELGELNIIAEDLGFTMDEVIELRRTG 359

QY 360 FPGMKILOFAPSGEDNAP-LDHNYPAGHNVVYVSGTHDNDTTLGFRTAPEAERAFMAY 418

Db 360 FPGMKILOFAPNPEDESIDSPHLAPA--NSWYITGTHDNNITVLGWTN--EIDDA-TREY 414

QY 419 LARYGIRCLSEYE-VAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFFPGLGDNWAW 477

Db 415 MARYTNR--KEYETVHAMLATVFSVSSFMALATWQDLELDEAARMNFFPGLGDNWAW 472

QY 478 YAEGLDFGLAAGLALAEASQR 500

Db 473 MTEDLTPAVEGLLDLTITVRR 495

RESULT 4

B98111

4-alpha-glucanotransferase (BC 2.4.1.25) [imported] - Streptococcus pneumoniae (strain

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004

C:Accession: B98111



A:Cross-references: UNIPROT:O66937; UNIPARC:UPI0000056423; GB:AE000704; NID:g2983301;  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: malM  
C:Superfamily: 4-alpha-glucanotransferase

Query Match 40.5%; Score 1100; DB 2; Length 485;  
Best Local Similarity 45.8%; Pred. No. 1.1e-79;  
Matches 219; Conservative 74; Mismatches 169; Indels 16; Gaps 6;

QY 5 RAFGILLHPTSPFGRWGIGALGREARERFLDWLADAGARWQVLPIGPTSY--GDSPYQSF 62  
DB 2 RLAGILHVTLSPPYIGDGLGKAYRFLDKCEGSLWQVLPNLPNTSLEAGNSYASN 61  
QY 63 SAFAGNPYLVDPEMLIISKGMLEQSEAPPPYQRTQVDYGMVYQTRWPLRLRRAPAGFRARAS 122  
DB 62 SLFAGNVYLDIPELLEEDLIKERDL-KRFFPLGEALYEVVYKELLEKAPKNFRR--- 117  
QY 123 AQDKTRLEAFTEARFMLEDYALFMALKTRFDGPGFWNWSPELDRDREPAALARAEELEAE 182  
DB 118 ---PELLEDFLKEHSYWLRYDALYMAIKEE--EGKEWYEWDEBELKRREKEALKRVLNKLGK 173  
QY 183 EVALYEWIOMLFYLEWQGTAKYAESKGTIIGDMPIFVAFDSSDVWNPQYFYLEADGNP 242  
DB 174 RFYFHVHVQVFFQWKEKRLYARERGISIVGDLPMYPSYSSADVWNPPELFLKDGDLKP 233  
QY 243 TVVAGVPRDYFSETGQLWGNPNLYRWDMERDNFAWCIARIQSLKQCHLVRIIDHFRGFEA 302  
DB 234 LFVAGVPPDFSKTGQLWGNPNVYVWEEHEKGGFRWIRRVHVNKLKDFLRLDHFGRFPA 293  
QY 303 YWEVPGFRPNVAGRWKAPGKLFPAVRAQLSDAPIAEDLGVTTPVEALRDGFGPFG 362  
DB 294 YWEPVYGEETAVNGRWKAPGKTLFKKLLSYFPKNPPIAEDLGFTTDBRVYLRTEFKIPG 353  
QY 363 MKILOAFSGEDNAFLPHNPYAHGNVVVYSTHNDITLGHWERTAPEARAERPMAYLARY 422  
DB 354 SRVIEAFYDKSEHPLHN--VEENNVYVYTHDLPPIRGWFMENLGEESKRKLFYLGRE 411  
QY 423 GIRCLSEYVAGALIELAFKSPAKLAIPLQDVGLGAPEARMPNPPGRIGDNWAWYAE 480  
DB 412 ---IKEEKNVESLIRLVLSRAKFAIIQODLLNLGNEARMYPPGRFPGNWRRIKE 465

RESULT 6  
A45049  
4-alpha-glucanotransferase (EC 2.4.1.25) - potato  
N:Alternate names: amylomaltase; D-enzyme; disproportionating enzyme  
C:Species: Solanum tuberosum (potato)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: A45049  
R:Takaha, T.; Yanase, M.; Okada, S.; Smith, S.M.  
J. Biol. Chem. 268, 1391-1396, 1993  
A:Title: Disproportionating enzyme (4-alpha-glucanotransferase; EC 2.4.1.25) of potato.  
A:Reference number: A45049; MUID:93123262; PMID:7678257  
A:Accession: A45049  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-576 <TAK>  
A:Cross-references: UNIPROT:Q06801; UNIPARC:UPI0000129746; GB:X68664; GB:SS2648; NID:g9;  
A:Experimental source: cv. May Queen, tubers  
A>Note: sequence extracted from NCBI backbone (NCBIN:122077, NCBIPI:122078)  
C:Superfamily: 4-alpha-glucanotransferase  
C:Keywords: chloroplast; glycosyltransferase; hexosyltransferase

Query Match 38.2%; Score 1035.5; DB 1; Length 576;  
Best Local Similarity 44.9%; Pred. No. 1.9e-74;  
Matches 216; Conservative 72; Mismatches 178; Indels 15; Gaps 6;

QY 4 ORAFGILLHPTSPFGRWGIGALGREARERFLDWLADAGARWQVLPIGPT---SVGDSPY 59  
DB 77 RRRAGILLHPTSPFPGYIGDLGPQAFKFDLWHLACSLWQVLPVPPGKRGHEDGSPY 136  
QY 60 QSFSAFAGNPYLVDPEMLIEKGMLEQSEAPPPYQRTQVDYGMVYQTRWPLRLRRAPAGFRA 119

137 SQDANCNTLLISLEELVDDGLKMEELPELPDTRVNYTISIEIDPLITKA-AKRL 195  
120 RASQDKTRLEAFIEAERF--WLEDYALFMALKTRFDGKPNWSPSLRDEPAALARAR 177  
196 SSEGELKDQLENFRDRPNISSWLEDAAYFAALDINSVNTISWYDWPPELKNRHLAALAEVY 255  
178 BELAEVALYEWIOWLFVLEWGTQKAYAESKGIQIIGDMPFVAFSDSDWANPOYFYLE 237  
256 QSEKDFIDIFTAQOFLFORQWKVRYARSKGISIMGDMPIYVYGHSAVWANKQFLLN 315  
238 ADGNPTVAGVPRDYFSETGOLGNPLRYDWMERDNFAWCIAIROSLKQCHLVRIIDHF 297  
316 RKGFPLVSGVPPDAFSETGOLGMSPLDYWKAMEKGFSSWVRRIQRAITDLDFFRDIHF 375  
298 RGFEAYMEVPPGRPNVAGRWKAPGEKLEFAAQAQLSDAPIIAEDLGVIPTPEVEALRDG 357  
376 RGFAFWAVPSEKTAIILGRKVGKPLFDAILQAVGKINIIEADLGVIPTPEVEALRDG 435  
358 FGFGMKILOPAP-SGEDNAPLPHNYPAHGNVYVSGTHDNDTLGFWRTAPAPAPMR 416  
436 IEAPGMVAVLQAFSGDAENPHLPHNH--EQNQVYVYTGTHDNDTLGFWRTAPAPMR 493  
417 AYLARYGIRCLSEVEVAGALIELAFKSPAKLAIVPLQDVLGPGPARNVPPGRLGDNWAW 476  
494 KYLSN-----IEEBEISRGLIEGAVSSVARIIIPMQDVLGSGDSRNIPATOFGNWSW 548  
477 R 477  
549 R 549  
RESULT 7  
F70120  
4-alpha-glucanotransferase (malO) homolog - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 09-Jul-2004  
C:Accession: F70120  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kervilavag, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: F70120  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-506 <KLB>  
A:Cross-references: UNIPROT:051188; UNIPARC:UPI0000057391; GB:AE001127; GB:AE000783; NID  
A:Experimental source: strain B31  
C:Superfamily: 4-alpha-glucanotransferase  
Query Match 26.8%; Score 727; DB 2; Length 506;  
Best Local Similarity 33.1%; Pred. No. 5.6e-50;  
Matches 173; Conservative 87; Mismatches 216; Indels 46; Gaps 8;  
QY 1 MOLORAFGILLHPTSPGRCWIGALGREAERFLDMLADAGARWQVPLGPTSYGDS-PY 59  
DB 10 LNLKRSKILLNISLPSKYIGDLGKAYKFDLFPASSQYQWQFAYSPIDFTRSPY 69  
QY 60 QSFSAFAGNPLYVDPEML-----ISKWLEQSEAPPYPTQRYDVGWLYQTRW----- 107  
DB 70 SIFSAGAGVYVIDEALDKFIDSLSLKENE-----TRYSLKKIS 112  
QY 108 ---PILRRAPAGFRARAAQDKTRLEAEERFWLEDYALFMALKTRFDGKPNWSPSLRDEPAALARAR 162  
DB 113 FKDKFLKEAALNINFRASADEVRSFKKSSYWLDDFAGVAFKFFKESKNFANVL 172  
QY 163 --PELRDRPALAREELAEVALYEWIOWLFVLEWGTQKAYAESKGIQIIGDMPFV 220  
DB 173 FDRGILKRNKDLFKRNILSKBIQVQLQYFFFSQFQALKRYANDKGIELIMVPPFI 232

221 AFSSDSDWANPOYFYLEADGNPTVAGVPRDYFSETGOLGNPLRYDWMERDNFAWCIA 280  
233 AYDSADWAWYOKYFKLRFDFASDKKIAGISPDYFLEQEQANDSAAYSNVLKFKFYEWAK 292  
281 RIRQSLKQCHLVRIIDHFRGFEAYWEVFGPRNPAVEGRWVKAPEGKLEFAAQAQLSDAPII 340  
293 RIGVLRKYADIIDHFRGFEVSTWESAGESYAFNGLWVKSPPGRDPFNFIILKIDKLKIW 352  
341 AEDLGVIPTPEVEALRDGFGPGKILQIFAPS-GEDNAPLPHNYPAHGNVYVSGTHDNDT 399  
353 VEDFONDLEVDLSRLFFNFPFGMKIMNLAFDFDSSNQNLPHNYK--NCIYVYTGSGDNDT 410  
400 TLGWERTAPAEARAFKRAYLARYGIRCLSEVEVAGALIELAFKSPAKLAIVPLQDVLG 459  
411 IREFINSDDLHKYIFDYLN-----TNEFVWDMIRSAMGVSVDNVISMQDIYILG 464  
460 PEARNMPPGRLGDNWAWYAEGLPEGLAAGLALAEASORA 501  
465 DKFSANIPKSTLDCNCFRLLSDLDATLSQNSIPITRLYGRA 506  
RESULT 8  
G72091  
4-alpha-glucanotransferase CP0431 [imported] - Chlamydia pneumoniae (strains CWL02;  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: G72091; F81577  
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lamell, C.; Fan, J.; Olinger, L.; Grimwood, J.  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606; PMID:10192388  
A:Accession: G72091  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-526 <ARN>  
A:Cross-references: UNIPROT:Q928L2; UNIPARC:UPI0000047BEF; GB:AE001617; GB:AE001363; N:  
A:Experimental source: strain CWL029  
R:Read T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255; PMID:10684935  
A:Accession: F81577  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-526 <REA>  
A:Cross-references: UNIPARC:UPI0000047BEF; GB:AE002204; GB:AE002161; NID:97189350; PDI  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: malQ, CP0431  
C:Superfamily: 4-alpha-glucanotransferase  
Query Match 16.6%; Score 451; DB 2; Length 526;  
Best Local Similarity 26.8%; Pred. No. 5.3e-28;  
Matches 131; Conservative 87; Mismatches 212; Indels 58; Gaps 15;  
QY 21 GIGALGREAERFLD-----WLADAGARWQVPLGPTSYGDSYQSFSAFAGNPLYVDP 74  
DB 41 GIG-----BFLDLPLISGCKQGFVQLPLNDTGEDTSPYNSISSVALNPLFLSL 93  
QY 75 EML-----IEKWLQSEAPPYPTQRYDVGWLYQTRWPLLRAPAGFRARASADK 126  
DB 94 SLEPNIDTIEVAKKLDQMHEL-----CSTPVSYSYQVKEKWAFLREYQKC-CRSSLSGN 149  
QY 127 TRLEAFIEARFWLEDYALFMALKTRFDGKPNWSPSLRDEPAALAREELAEVAL 186  
DB 150 SNFSEFLESERYWLYPGTFFRAIKHMHGEPINNPKSLTDQE--NPPDLTKKFHDEVL 207  
QY 187 YEWIOWLFVLEWGTQKAYAESKGIQIIGDMPFVAFSDSDWANPOYFYLEADGNPTV 246  
DB 208 FSYLQFLCYOOLCEVKAYADQHVLLKGDPLILSKDSCDWWYFRDYF-----SSRSV 261  
QY 247 GVPDYFSETGOLGNPLRYDWMERDNFAWCIAIROSLKQCHLVRIIDHFRGFEAYWEV 306

Db 262 GAPFDLYNSEGQWHLPIYFNSQLAKDDYIWWKERLYAQNFYSVYRLDHLIIGFFRLWI - 320  
QY 307 PFGPNVAGEGRWVK-----APEGKLPAAVRAQLSDAPIIAEDLGWITVEVALRDGFG 359  
Db 321 ---WDSGRCGRFIPDPKDYIKQGTETLSTMLGASSMLP-IGEDLGIIPOQDVKTTLTHLG 376  
QY 360 FPGMKILOQAFSGE-DNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFTAPAEARAFMAY 418  
Db 377 ICGTRIPRWERNWESDGAFLPKDYNPLSVTLTSTHSDTFAQWNLNSPKAEAKQAFKL 435  
QY 419 LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPE-----ARNMPPGRL 470  
Db 436 HLPFOKTLTTETQI--DILKLSHESASIFHINLNDVLCPLDVLSKNLQORINTPTGI 493  
QY 471 G-DNNAWR 477  
Db 494 SKNWSYR 501

RESULT 9  
F86531  
glucanotransferase [imported] - Chlamydomophila pneumoniae (strain J138)  
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: F86531  
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ito, Y.  
Nucleic Acids Res. 28, 2311-2314, 2000  
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A;Reference number: A86491; MUID:20330349; PMID:10871362  
A;Accession: F86531  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-526 <STO>  
A;Cross-references: UNIPROT:Q928L2; UNIPARC:UPI0000047BEF; GB:BA000008; NID:98978700; PT  
A;Experimental source: strain J138  
C;Genetics:  
A;Gene: malQ.  
C;Superfamily: 4-alpha-glucanotransferase

Query Match 16.6%; Score 451; DB 2; Length 526;  
Best Local Similarity 26.8%; Pred. No. 5.3e-28;  
Matches 131; Conservative 87; Mismatches 212; Indels 58; Gaps 15;

QY 21 GIGALGREAERFLD-----WLADAGARWQVPLGPTSYGDSYQSFAGNPNYLVDP 74  
Db 41 GIG-----EFLDLPLISWCQKQGSFVQLPLNDTGEDTSPYNSISSVALNPLFLSL 93  
QY 75 EML-----IEKGWLEQSEAPPYPTQRTVDYGMVLYQTRWPLLRRAFAFRARASADK 126  
Db 94 SSLPNIDTIVEVAKKQDMHEL---CSTPSVSYTVQVKEKWAFLUREYQK-CKSSLEGN 149  
QY 127 TRLEAFTEABRFWLEDYALFMALKTRDPGKPNWSPPELRDREPAALAREELAEVAL 186  
Db 150 SNSEFLESERYMLPYGTFRANKHMHGEPINNPKSLTDQE--NFPDLTKKFHDEVLF 207  
QY 187 YEWIQLFYLEWGTAKYASKGIIIGDMPIFVAFDSSDVWANPQVLYLEADGNPTVA 246  
Db 208 FSYLQFLCYOOLCEVKAQDQHHVLLKGLDPLILISKDCSDVWYPRDYF-----SSRSV 261  
QY 247 GVPDIFYSETGOLWGNPLXWMDNERDNFACIARIQSLKQCHLVRIIDHFRGFEAYEV 306  
Db 262 GAPFDLYNSEGQWHLPIYFNSQLAKDDYIWWKERLYAQNFYSVYRLDHLIIGFFRLWI - 320  
QY 307 PFGPNVAGEGRWVK-----APEGKLPAAVRAQLSDAPIIAEDLGWITVEVALRDGFG 359  
Db 321 ---WDSGRCGRFIPDPKDYIKQGTETLSTMLGASSMLP-IGEDLGIIPOQDVKTTLTHLG 376  
QY 360 FPGMKILOQAFSGE-DNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFTAPAEARAFMAY 418  
Db 377 ICGTRIPRWERNWESDGAFLPKDYNPLSVTLTSTHSDTFAQWNLNSPKAEAKQAFKL 435  
QY 419 LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPE-----ARNMPPGRL 470

Db 436 HLPFOKTLTTETQI--DILKLSHESASIFHINLNDVLCPLDVLSKNLQORINTPTGI 493  
QY 471 G-DNNAWR 477  
Db 494 SKNWSYR 501

RESULT 10  
G71557  
probable glucanotransferase - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C;Species: Chlamydia trachomatis  
C;Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 09-Jul-2004  
C;Accession: G71557  
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell  
Science 282, 754-759, 1998  
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tr  
A;Reference number: A71570; MUID:99000809; PMID:9784136  
A;Accession: G71557  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-527 <ARN>  
A;Cross-references: UNIPROT:O84089; UNIPARC:UPI0000047C0C; GB:AE001283; GB:AE001273; NJ  
A;Experimental source: serotype D, strain UW-3/Cx  
C;Genetics:  
A;Gene: malQ  
C;Superfamily: 4-alpha-glucanotransferase

Query Match 15.9%; Score 430.5; DB 2; Length 527;  
Best Local Similarity 25.9%; Pred. No. 2.3e-26;  
Matches 126; Conservative 89; Mismatches 233; Indels 39; Gaps 12;

QY 33 LDWLADAGARWQVPLGPTSYGDSYQSFAGNPNYLVDPPEMLIEKGLQSEAPPY 92  
Db 56 IDMCISGFOILQILPINDTGSCSPYNSISSALNPLHLSISALPYKEEVPAAETIRE 115  
QY 93 PTQRVDYGMVLYQTRWPLLRRAFAFRARASAQ---DKTRLEAFTEABRFWLEDYALFMA 148  
Db 116 MQQLSQLQVHYEKVRSKMKEDFFQYYRVCKQKLLDHPDPYAFCEQEKYWLHFLYFRS 175  
QY 149 LKTRFDGKPNWSPPELRDREPAALAREELAEVALYEWIQLFYLEWGTAKYAESK 208  
Db 176 IREHLONLPIINHWPPTTYTDL--SQITEHERTFADIQFHSYLYQLYCFQOQMTVREHANCK 233  
QY 209 GIIQIGDMPIFVAFDSSDVWANPQVLYLEADGNPTVAVGPRDYFSETGOLWGNPLYRWD 268  
Db 234 SCLIKGDIPLILISKDCSDVWYPRDYF-----SSSESVGAPPDLYNAEQNWHLPICNMK 287  
QY 269 VMERDNFACIARIQSLKQCHLVRIIDHFRGFEAYW---EVPEGRPNVAGEGRWVKAPEGK 325  
Db 288 TLQODNYLWKEKRLRYAENFYSLYRLDHHVVLGFRFWWDESGCGRPEPHDPKNYLAQGD 347  
QY 326 LFAAARAQLSDAPIIAEDLGWITVEVALRDGFGFGMKILOQAFSGEDN-AFLPHNYPA 384  
Db 348 ILSHLTSSSMLP-IGEDLGTPSDVVKRMLESFAVCGTRIPRWERNWEGNAYTP--FDQ 404  
QY 385 HGNVVVYS-GTHDNDTTLGWFTAPAEARAFMR---AYLARYGIRCLSEYEVAGALIEL 439  
Db 405 YDPLSVTSLTSDSSSLASWWSKESQESKLAFAQLGLFYSSTLSLHNHTE-----ILKL 458  
QY 440 AFKSPAKLAIVPLQDVLGLGPE-----ARNMPPGRLG-DNNAWRYABGDLPEPGLAAG 490  
Db 459 SHKTSSIFRINLINDYLALEFPDLISKTPRYERINLFGTISKNNVYRVKVSIEDLSHSK 518  
QY 491 LRALAEA 497  
Db 519 LNSLLEA 525

RESULT 11  
A81712  
4-alpha-glucanotransferase TC0362 [imported] - Chlamydia muridarum (strain Nigg)  
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C:Superfamily: 4-alpha-glucanotransferase  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 12.6%; Score 341.5; DB 2; Length 698;  
Best Local Similarity 26.0%; Pred. No. 4e-19;  
Matches 140; Conservative 76; Mismatches 245; Indels 77; Gaps 23;

QY 20 WIGGALGREAEFLDWLAGARWVCLP---LGPTS-YGDSYPQSFSAFAGNPVLYDPE 75  
DB 162 WIGGPGF-DLKQMLEQVGERGSGFGLNFIHALYPANPHSASPSRSSRLNVIYDIN 220  
QY 76 MLIE-----KGWLEQSEAPPYPTQR-----VDYGMLYQTRWPLLRARFAGFPARASQA 124  
DB 221 RVEEFQOSEAQRWQHQAETQALAKRSSEWVDYPLVMQLKLTALRLSPFLFTARKAKD 280  
QY 125 DKTR-LEAFIEAEFRWLEDYALFMALKTRF-DGKP-----WNWSPELDRREPALARARE 178  
DB 281 AQVQAFRHFVEQGGSLHQQAVFDALHAHLSEHDPMMWGPVWPEKRYDRGHSSAVADFGR 340  
QY 179 ELAEVALYEWLQWLFYLSWGQTKAVAESK-GIIGDMPIFVAFDSSDVWANPQVFFYL 236  
DB 341 ERADEFTVILNQLWLAASQFDDCFQSQQRKMPGIGHYRLDLAVGAEQGAETWCDELRYCL 400  
QY 237 EADGNFTVAVGPRDYFSETGQWGNPLXYRWDMYERDNFAWCIARIQSLKQCHLVRIDH 296  
DB 401 KAS-----VGAPPDILPLGQNWGLPMPDPHVVARAYQPFIDLLRANMTSCGALRIDH 454  
QY 297 FRGFEAYEVPGRPNVAGRWVKAGEKLF--AVRAQLSDAPIAIEDLGVITPE-VEA 353  
DB 455 VMLALRLMWIPYGH-TADQGAIVKYVVDLLVALLESQRHCHMVGIEDLGTVPVEIVGK 513  
QY 354 LRDFGFGPKILQPAFSGEDNAFLPHNYPAHGNVVYSGTHDNTDTLGNWFT----- 406  
DB 514 LRDS-GVYSKYVLYFHDSENIIFRAQSYVQAMATI--THDLPTLRGYQWADDTLGN 570  
QY 407 ----APEAEARFMRAYL----ARYG-IRCLSEYEV----AGALIELAFKSPA-----KL 447  
DB 571 KLGLYPD-QQILKQLYLDRERAKQGLLEGHLRYDVCVKVGHKAALLSMSPLNRGLQRY 629  
QY 448 AIVPLQDVILGPE-----ARNVFPRLGNWAWRYAEGDLEPLAAGLRALAEASQ 499  
DB 630 VADSASALIGLOPEDWLDMAAPYNIPTTDEYPNWRK-----LSASLEEIPADSQ 680

RESULT 13  
G70928  
probable 4-ALPHA-GLUCANOTRANSFERASE - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C:Accession: G70928  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, N.;  
Rajandream, M.A.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: G70928  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1724 <COL>  
A:Cross-references: UNIPROT:O53932; UNIPARC:UPI000012EB5C; GB:AL022021; GB:AL123456; N1  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv1781c  
C:Superfamily: 4-alpha-glucanotransferase

Query Match 11.8%; Score 320.5; DB 2; Length 724;  
Best Local Similarity 27.2%; Pred. No. 2e-17;  
Matches 153; Conservative 69; Mismatches 258; Indels 83; Gaps 26;

QY 4 QRAFGILLHTTSPGR--WGIGALGREAEFLDWLADA-GARWQVLPPL-----GPTSY 54

172 RAWGLAVOLYXVRSSQSGWIGDITDLANLAL-WSASAHGAGYVLVNLPHAAATLPGPAGR 230  
55 G-----DSYQSFSAFAGNP-VL---VDPEM--LIEKGWLEQSEAPPPYPTQRYDYGWLYQ 104  
231 SKPIEPSYLPSTRFRFVNPFLYRVEAIPELVDLPKGRGVQRLRTNVVQAHADOLD-TIDRD 289  
105 TRWPLLRRAAF-AGFRARASQDKTRLEAFIEAERFWLEDYALFMAKTRFPDCKPWNWSP 163  
290 SAMAARAAKLIVHVRPSAGRELAYAAFRITREGALDDFATWALAEY-GDDWHRWPK 348  
164 ELRDREPAALARARELAEVALYEWIOWLF--YLEWQTKAYAESKGIQIIGDMPFVA 221  
349 SLRHPDASGVADFVDKHADAVDFHRLQWQDLQELASQSAQALRAGMSLGMADLAVGVH 408  
222 FDSDDVWNPQFYFYLEADGNPTVAGVPRDYFSEGTQWLGWGNPLRYWDMVMDRDNFAWCJAR 281  
409 PNGADAWALQDVL---AQS---VTAGAPPDEFNQLGODMSPWPWRPDLASQERYPPFAL 462  
282 IROSLKQCHLVRIIDHFRGFEAYWVPPGRPNVAGRWYKAPCEKL--FAAVRAQLSDAPI 339  
463 IQAALRHAGAVRIDHIIIGLFLRWIIPDGAP-PTQGTYYRYDHDAIGIVALEAHAGVAV 521  
340 IAEGLVITTEVEALRDGF---GPPGMKILQAFSGEDNAPLPHNYPAHGN-----VVVY 391  
522 VGEDLTGVEPVW---RDYLLRLGLLGTSLWFE---QDRDCGPAGTPLPAERWREYCLSS 575  
392 SGTINDNTLWFR-----TAP-----EASRPFRYLAARYGIRC----- 426  
576 VTTHDLPPTAGLAGDQVRLRESGLLTPNVEALESARADRAAMAELELRVGLADGAE 635  
427 LSEYEVAGALTELAFKSPAKLAIPLQDVLGLGPEARMNFPRLGD--NNAWRYAEGDLE 484  
636 PDSEAVLALRYLGRTPSRLLAVALTD--AVGDRTONQFGTTDEYPNWRVPLTGPDCQ 693  
485 PGLAAGL-----RALAEASORA 501  
694 PMLLEDIFDRAATLAEAVRAA 716  
RESULT 14  
F83375  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A:Reference number: AB2950; MUID:2043737; PMID:10984043  
A:Accession: F83375  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-684 <STO>  
A:Cross-references: UNIPROT:Q911V2; UNIPARC:UPI000000C55E0; GB:AE004643; GB:AE004091; NID  
A:Experimental source: strain PA01  
C:Genetics:  
C:Superfamily: 4-alpha-glucanotransferase  
Query Match 11.6%; Score 313.5; DB 2; Length 684;  
Best Local Similarity 25.2%; Pred. No. 6.6e-17;  
Matches 145; Conservative 70; Mismatches 258; Indels 103; Gaps 23;  
4 ORAFGILLHPTSPFG-----RWGI-----GALGREA--ERFLDWLADA 39  
125 QREWAVNAPSCSLASLAFGRGRWGLAAQVALRRPFGGGLGDSNALEDLRSARH 184  
40 GARWQVWLPL----GPTSYGDSYQSFSAFAGNPYLVDPEMLIEKGWLEQ-----SE 87  
185 GADALATISPLHALAEANGHAYSPSPSRSLFFNVLHAAPATILGNAAVEQAIRAGLAAE 244

88 APPPYPTQRYDYGWLYQTRWPLLRRAFPAGFRARASQDKTRLEAFIEAERFWLEDYALFM 147  
245 MARLESLELDWTAAADLRWELLRLQRLHRDFTERRASPLRHLDAEPREAGEALLH-HCRFE 303  
148 ALKTRFDGKP-WNEW-SPELRDRREPAALARARELAEVALYEWIOWLFY--LEWQTKA 203  
304 TLOAHLGAGDPDRWRPPEFLRRPGEPAVAACADH-AEEVDFFRAFQGLWLTQRCLOHAORQA 362  
204 YAESKGIQIIGDMPFI FVAFDSSDVWNPQFYFYLEADGNPTVAGVPRDYFSEGTQWLGWGNP 263  
363 REAGMALGVADLAVGADGGGSAWSQEBELLAE-----VNVGAPDILNOSQODWGV 416  
264 LYRWDVMDERDNFACIARIQSLKQCHLVRIIDHFRGFEAYWVPPGRPNVAGRWYKAPG 323  
417 AFNPEGLRRHGYAFREMLRANLANPCGLRIDHVMGLQRLWLI PRGQPPHA-GAYLRYPQ 475  
324 EKL--FAAVRAQSLSDAPIIAEDLGIVITTEVEALRDGF---FPGMKILQAFSGEDNAPL 378  
476 RELRLLEASASALVIGEDLGTVP---EGLREELARRQVLGTRVLLFPERRGE--RFV 530  
379 -PHNYPAGHGVVYVSGTHDNDTTLGWERTAPEAERAFMAYLARYGIR---CLSEYEA 433  
531 PPAQMPA--DAMATTSTHDLPSLSGWRGRDIHWR-----GRAGHRSABECAAULBLR 581  
434 G-----ALIELAFKSPAKLAIPLQDVLGLGPEARMNFP 468  
582 AEERRALAASLEPPVDPDAAEVPLDACIGVVGATPAPLVLLPLEDALGSLQPNLP 641  
469 RLGNWNRVYAEGLDLE-----PGLAAGLRALAEASQR 500  
642 DAHPNWRNRPPENAAQMLGTQVQDRLRL-LLESR 676  
RESULT 15  
AI0996  
A:Title: Complete genome sequence of *Salmonella enterica* subsp. *enterica* serovar Typhi  
A:Note: this species has also been called *Salmonella typhi*  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AI0996  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar  
S.; Moule, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.  
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* ser  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AI0996  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-693 <PAR>  
A:Cross-references: UNIPARC:UPI000000SA7AC; GB:AL513382; PIDN:CAD08101.1; PID:gl6505080.  
C:Genetics:  
C:Gene: malQ  
C:Superfamily: 4-alpha-glucanotransferase  
C:Keywords: Glycosyltransferase; hexosyltransferase  
Query Match 11.1%; Score 302.5; DB 2; Length 693;  
Best Local Similarity 24.7%; Pred. No. 5e-16;  
Matches 135; Conservative 68; Mismatches 238; Indels 105; Gaps 22;  
20 WGTGALGREAFELDWLADAGA-----RWQVPLGTPSYGDSYQSFSAFAGNPYLVD 74  
157 WGTGDFGLKSMVLVDVATRGGAFIGLNPILHPIALIPANPES--ASYSFSSRRWLVIIYDV 214  
75 EMLIEKGWLEQSEAPPPY-----TQRVDYGLWLYQTRWPLLRRAFPAGFRARASA 123  
215 NAVEDFRLSEEAQWQMPATQCKLRQARDAQWVDYATVTATKITALRMAWTRFAARDDA 274  
124 QDKTRLEAFTEAE---RFWLEDYALFMALKTRFDGKPNW--EWSPELRDRREPAALARE 178  
275 Q-MAEFRRHFTAREGESLYMQAFDALHAYQVKEDGQWGWPAWPEAYQSVESPAVKQFCE 333



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2006, 14:03:37 ; Search time 164 Seconds  
(without alignments)  
2155.303 Million cell updates/sec

Title: US-10-618-976-2

Perfect score: 2714

Sequence: 1 MQLQARFGLLHPTFFGRW.....DLRPLAAGLALAEASQRA 501

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1696.5	62.5	500	1	MALQ_THETH
2	1692.5	62.4	500	2	Q5SIV3_THERM8
3	1692.5	62.4	500	2	Q72J82_THERM2
4	1635.5	60.3	500	2	Q6JHX9_THEAQ
5	1626.5	59.9	500	2	Q6JHY0_THESC
6	1475	54.3	503	2	Q4HA16_9DEIO
7	1318	48.6	499	2	Q5N4Q2_SYNP6
8	1305.5	48.1	502	2	Q8YQG0_ANASP
9	1248	46.0	504	2	Q7NGS7_GLOVI
10	1245	45.9	493	2	Q74DY3_GEOSL
11	1222.5	45.0	505	1	MALQ_SYNY3
12	1221	45.0	518	2	Q4NJB1_9DELT
13	1217	44.8	518	2	Q8DK21_SYNEL
14	1194	44.0	505	1	MALQ_STRPN
15	1194	44.0	505	1	MALQ_STRR6
16	1174	43.3	497	2	Q8XHV6_CLOPE
17	1154	42.5	505	2	Q7UT73_RHOBA
18	1117.5	41.2	494	2	Q60817_METCA
19	1108.5	40.8	487	1	MALQ_CLOBU
20	1108.5	40.8	540	2	Q73JN0_TREDE
21	1100	40.5	485	1	MALQ_AQUAE
22	1056	38.9	497	2	Q8P0K6_STRP8
23	1055	38.9	497	2	Q5XBP8_STRP6
24	1041	38.4	497	2	Q992C2_STRP3
25	1039	38.3	497	2	Q8K751_STRP3
26	1035.5	38.2	576	1	DPEP_SQITU
27	1024.5	37.7	503	2	Q726G0_DESVH
28	1023.5	37.7	499	2	Q5LZT4_STRT1
29	1022.5	37.7	502	2	Q5M4E4_STRT2
30	1020	37.6	498	2	Q8DYN6_STRAS
31	1020	37.6	498	2	Q8E494_STRAS3

32	1016	37.4	601	2	Q6PYV7_QSTTA
33	1009	37.2	556	2	Q5V0X8_HALMA
34	1007.5	37.1	576	2	Q9LV91_ARATH
35	1006.5	37.1	468	2	Q8ZXM0_PYRAE
36	1001.5	36.9	493	2	Q7P7Q5_FUSNV
37	986	36.3	585	2	Q9FDV9_CHLRE
38	982.5	36.2	509	2	Q8DT30_STRMU
39	976.5	36.0	506	2	Q8RF60_FUSNN
40	923.5	34.0	586	2	Q8LI30_ORYSA
41	875.5	32.3	487	2	Q5NHN0_FRATT
42	841	31.0	516	2	Q7VBH6_PROMA
43	839.5	30.9	512	2	Q7U7L9_SYNFX
44	825	30.4	521	2	Q7V6S5_PROMM
45	796.5	29.3	506	2	Q7V110_PROMP

## ALIGNMENTS

### RESULT 1

ID	MALQ_THETH	STANDARD	PRT	500 AA
AC	O87172;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	10-MAY-2005 (Rel. 47, Last annotation update)			
DE	4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)			
DE	(disproportionating enzyme) (D-enzyme).			
GN	Name=maloQ;			
OS	Thermus thermophilus.			
OC	Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;			
OC	Thermus.			
OX	NCBI_TaxID=274;			
RP	[1]			
RP	NUCLEOTIDE SEQUENCE.			
RC	STRAIN=ATCC 33923 / DSM 674 / AT-62;			
RA	Terada Y., Fujii K., Takaha T., Okada S.;			
RA	"Cloning, expression and characterization of amylomaltase from Thermus aquaticus ATCC33923."			
RT	aquaticus ATCC33923."			
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.			
RP	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).			
RX	MEDLINE=20143895; PubMed=10677288; DOI=10.1006/jmbi.1999.3503;			
RA	Przydas I., Tomoo K., Terada Y., Takaha T., Fujii K., Saenger W.,			
RA	Strater N.;			
RT	"Crystal structure of amylomaltase from Thermus aquaticus, a glycosyltransferase catalysing the production of large cyclic glucans."			
RL	J. Mol. Biol. 296:873-886(2000).			
CC	-I- CATALYTIC ACTIVITY: Transfers a segment of a 1,4-alpha-D-glucan to a new position in an acceptor, which may be glucose or a 1,4-alpha-D-glucan.			
CC	-I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-I- SIMILARITY: Belongs to the disproportionating enzyme family.			
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.			
DR	EMBL; AB016244; BAA33728.1; -; Genomic_DNA.			
DR	PDB; 1CWY; X-ray; A=1-500.			
DR	PDB; 1ESW; X-ray; A=1-500.			
DR	PDB; 1FP8; X-ray; A=1-500.			
DR	PDB; 1FP9; X-ray; A=1-500.			
DR	InterPro; IPR003385; Glyco_hydro_77.			
DR	Pfam; PF02446; Glyco_hydro_77; 1.			
DR	TIGRFAMs; TIGR00217; maloQ_1.			
KW	3D-Structure; Carbohydrate metabolism; Glycosyltransferase;			
KW	Transferase.			
FT	STRAND	6	10	
FT	HELIX	13	15	





Query Match 62.4%; Score 1692.5; DB 2; Length 500;  
 Best Local Similarity 62.7%; Pred. No. 1e-119;  
 Matches 314; Conservative 66; Mismatches 118; Indels 3; Gaps 2;

QY 1 MOLQAFGLLHPTSPFGRWGIGALGREAEFLDLWADAGARWQVLPGLPTSYGDSYQ 60  
 DB 1 MELPRAFGILLHPTSLPGPYGVGLQEARDFLKEAGGRYQVLPGLPTGYGDSYQ 60

QY 61 SFSFAGNPLYVDPEMLIEKWLQSEAPPYPTQRTVDYGLYQTRWPLLRAPAFGRAR 120  
 DB 61 SFSFAGNPLYLDLRLAERGVRLED--PGPQGRVDYGLLYAKWPKALKEAFGRFKEK 118

QY 121 ASADKTRLEAFIEAEFRWLELYALFMALKTRFDGKPNWNSPELRDREPAALAREBEL 180  
 DB 119 ASPEERAEFAAFREAEAWLELYALFMALKGAHGLPNNRWPLPLRKEEKALREKASAL 178

QY 181 ABEVALYEWIQLFYLEWQTKAYAESKGIQIIGDMPFVAFDSDVWNPQYFYLEADG 240  
 DB 179 ABEVAFHAFQWLFFRQWGLKAEAEALGIRIIGDMPFVAFDSDVWNPQYFYLEADG 238

QY 241 NPTVAGVPRDYFSETGOLWGNPLYRWDMERDNFACIARIOSLKQCHLVRIIDHFRGF 300  
 DB 239 RPTVAVGPPDYFSETGQWGNPLYRWDLEREGFSFWIRLEKALEFLHLVRIDHFRGF 298

QY 301 EAYVEVPGRPNVAGRWKAPGKELFAAQAQLSDAPIAEIDLGVITPEVEALRDGFGF 360  
 DB 299 EAYWEIPASCTAVEGRWKAPGKELFKIQEVFGEVPLAEIDLGVITPEVEALRDGFGF 358

QY 361 PGMKILOPAF-SGEDNAFLPHNYPAHGNVYVSGTHDNTTLGWFRTAPEAEAFMAYL 419  
 DB 359 PGMKVLOPAFDGDMENPFLPHNYPAHGRVYVYTGTHDNTTLGWYRTATPEKAFMAYL 418

QY 420 ARYGIRCLSEYEVAGALIELAFKSPAKLAIUPLQVLGLGPEARNWPFGRGLDGNWARYA 479  
 DB 419 ADMGITTREBEVFPWALHLMGKSVARLAVYVQDVLALGSEARNWYVGRPSGNWAWRL 478

QY 480 EGDLEPGLAAGLALAEASOR 500  
 DB 479 PGELSPHAGRLAWAEATER 499

RESULT 3

QY Q72J82\_THET2 PRELIMINARY; PRT; 500 AA.  
 AC Q72J82;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE 4-alpha-glucanotransferase (EC 2.4.1.25).  
 GN OrderedLocusNames=trc0897;  
 OS Thermus thermophilus (strain HB27 / ATCC BAA-163 / DSM 7039).  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
 OC Thermus.  
 CX NCBI\_TaxID=262724;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed=15064768; DOI=10.1038/nbt956;  
 RA Henne A., Brueggemann H., Raasch C., Wieser A., Hartsch T.,  
 RA Liesegang H., Johann A., Lienard T., Gohl O., Martinez-Arias R.,  
 RA Jacobi C., Starkuviene V., Schlenczek S., Dencker S., Huber R.,  
 RA Klenk H.-P., Kramer W., Merkl R., Gottschalk G., Fritz H.-J.;  
 RT "The genome sequence of the extreme thermophile Thermus  
 thermophilus.";  
 RT Nat. Biotechnol. 22:547-553 (2004).  
 RL EMBL; AE017304; AAS81241.1; -; Genomic\_DNA.  
 DR SMR; Q72J82; 1-500.  
 DR GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR003385; Glyco\_hydro\_77.  
 DR Pfam; PF02446; Glyco\_hydro\_77; 1.  
 DR TIGRFAMs; TIGR00217; malQ; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 500 AA; 57194 MW; C4B5A3EDD9DC1336 CRC64;

Query Match 62.4%; Score 1692.5; DB 2; Length 500;  
 Best Local Similarity 62.7%; Pred. No. 1e-119;  
 Matches 314; Conservative 66; Mismatches 118; Indels 3; Gaps 2;

QY 1 MOLQAFGLLHPTSPFGRWGIGALGREAEFLDLWADAGARWQVLPGLPTSYGDSYQ 60  
 DB 1 MELPRAFGILLHPTSLPGPYGVGLQEARDFLKEAGGRYQVLPGLPTGYGDSYQ 60

QY 61 SFSFAGNPLYVDPEMLIEKWLQSEAPPYPTQRTVDYGLYQTRWPLLRAPAFGRAR 120  
 DB 61 SFSFAGNPLYLDLRLAERGVRLED--PGPQGRVDYGLLYAKWPKALKEAFGRFKEK 118

QY 121 ASADKTRLEAFIEAEFRWLELYALFMALKTRFDGKPNWNSPELRDREPAALAREBEL 180  
 DB 119 ASPEERAEFAAFREAEAWLELYALFMALKGAHGLPNNRWPLPLRKEEKALREKASAL 178

QY 181 ABEVALYEWIQLFYLEWQTKAYAESKGIQIIGDMPFVAFDSDVWNPQYFYLEADG 240  
 DB 179 ABEVAFHAFQWLFFRQWGLKAEAEALGIRIIGDMPFVAFDSDVWNPQYFYLEADG 238

QY 241 NPTVAGVPRDYFSETGOLWGNPLYRWDMERDNFACIARIOSLKQCHLVRIIDHFRGF 300  
 DB 239 RPTVAVGPPDYFSETGQWGNPLYRWDLEREGFSFWIRLEKALEFLHLVRIDHFRGF 298

QY 301 EAYVEVPGRPNVAGRWKAPGKELFAAQAQLSDAPIAEIDLGVITPEVEALRDGFGF 360  
 DB 299 EAYWEIPASCTAVEGRWKAPGKELFKIQEVFGEVPLAEIDLGVITPEVEALRDGFGF 358

QY 361 PGMKILOPAF-SGEDNAFLPHNYPAHGNVYVSGTHDNTTLGWFRTAPEAEAFMAYL 419  
 DB 359 PGMKVLOPAFDGDMENPFLPHNYPAHGRVYVYTGTHDNTTLGWYRTATPEKAFMAYL 418

QY 420 ARYGIRCLSEYEVAGALIELAFKSPAKLAIUPLQVLGLGPEARNWPFGRGLDGNWARYA 479  
 DB 419 ADMGITTREBEVFPWALHLMGKSVARLAVYVQDVLALGSEARNWYVGRPSGNWAWRL 478

QY 480 EGDLEPGLAAGLALAEASOR 500  
 DB 479 PGELSPHAGRLAWAEATER 499

RESULT 4

QY Q6JHX9\_THEAQ PRELIMINARY; PRT; 500 AA.  
 AC Q6JHX9;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE 4-alpha-glucanotransferase.  
 OS Thermus aquaticus.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
 OC Thermus.  
 CX NCBI\_TaxID=271;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Park K.-H., Park J.-H.;  
 RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY459352; AAR23242.1; -; Genomic\_DNA.  
 DR SMR; Q6JHX9; 1-499.  
 DR GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR003385; Glyco\_hydro\_77.  
 DR Pfam; PF02446; Glyco\_hydro\_77; 1.  
 DR TIGRFAMs; TIGR00217; malQ; 1.  
 KW Transferase  
 SQ SEQUENCE 500 AA; 56973 MW; A03680E3B3113256 CRC64;

Query Match 60.3%; Score 1635.5; DB 2; Length 500;  
 Best Local Similarity 61.7%; Pred. No. 2.2e-115;  
 Matches 309; Conservative 59; Mismatches 130; Indels 3; Gaps 2;

```

QY 1 MOLQAFGLILLHPTSPGQWIGALGREAREFLDMLADAGARWQVPLGPTSYGDSYQ 60
DB 1 MELPRAFGILLHPTSLPGPYGVGLGLEARDFLRLKAGGGRFQVPLGPTGYGDSYQ 60
QY 61 SFSFAGNPNLYDPEMLTEKGLWLEQSEAPPYPTQRYDYGWLYQTRWPLLRARAFAGFRAR 120
DB 61 AFSFAGNPNLYDPLRLEKGLVLUKD--PGFQGRVDYGYWLYAKWYKPAALKAAYQGFLEK 118
QY 121 ASADKTRLEAFIEAERFWLEDYALFMAKTRFDGKPNWSPPELRDRPAALAREEL 180
DB 119 APRKEREDFPAQOEKESWLDYALFMAKQAQHGGLPWNRPPLRRREKAMKEAEAL 178
QY 181 ABEVALYEWIOWLFYLENGQTKAYAESKGIQIIGDMPFVAFDSSDVWNPQYFVLEADG 240
DB 179 ABEVAFHAWTOWLFPEAWKALKEEAALGIIQIGDMPFVAFDSSAEVWNPQYFVLEADG 238
QY 241 NPTVAGVPRDYFSETGOLWGNPNLYRWDVMDERDFANCIARIQSLKQCHLVRIIDHPRGF 300
DB 239 RPLVAGVPPDYFSETGOLWGNPNLYRWDVMDERDFANCIARIQSLKQCHLVRIIDHPRGF 298
QY 301 EAYWEVPPGRPNNAVGRWVKAPGEKLFPAARQAQSLDAPIIAEDLGVIITPEVEALRDGFGF 360
DB 299 EAYWEIPASCTAVGRWVKASGEKLFDRIOEVFGEVPIAEDLGVIITPEVEALRDGFGF 358
QY 361 PGMKILOFAPS-GEONAFPLPHNYPAHGNVYVSGTHDNDTTLGWFTAPAEARAFMAYL 419
DB 359 PGMKVLOFAPDHGMENPFLPHNYPAHGNVYVSGTHDNDTTLGWFTAPAEARAFMAYL 418
QY 420 ARYGIRCLSEVEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRGLDNWAWRYA 479
DB 419 ADMGISFQEEVFWALMGLCLKSVARLAIYPVQDVLALGSEARMNYPGRSGNWAWEALL 478
QY 480 EGDLEPGLAAGLALAEASOR 500
DB 479 PGQLTQEHARLAWAEATGR 499

RESULT 5
Q6JHYO_THESC PRELIMINARY; PRT; 500 AA.
AC Q6JHYO_THESC
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE 4-alpha-glucanotransferase.
OS Thermus scotoductus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=37636;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Park K.-H., Park J.-H.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY459351; AAR23241.1; -; Genomic_DNA.
DR SMR; Q6JHYO; 1-500.
DR GO; GO:0004134; P:4-alpha-glucanotransferase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR003395; Glyco_hydro_77.
DR Pfam; PF02446; Glyco_hydro_77; 1.
DR TIGRFAMs; TIGR00217; malQ; 1.
KW Transferase.
SQ SEQUENCE 500 AA; 57730 MW; 876B2DA81F61B33A CRC64;

```

```

Query Match 59.9%; Score 1626.5; DB 2; Length 500;
Best Local Similarity 60.3%; Pred. NO. 1.1e-114;
Matches 302; Conservative 66; Mismatches 130; Indels 3; Gaps 2;

QY 1 MOLQAFGLILLHPTSPGQWIGALGREAREFLDMLADAGARWQVPLGPTSYGDSYQ 60
DB 1 MELPRAFGILLHPTSLPGPYGVGLGLEARDFLRLKAGGGRFQVPLGPTGYGDSYQ 60
QY 61 SFSFAGNPNLYDPEMLTEKGLWLEQSEAPPYPTQRYDYGWLYQTRWPLLRARAFAGFRAR 120

```

```

DB 61 SFSFAGNPNLYDPLRRLRRRGYLRED--PGFEGRVVYGYWLYAKWYKPAALAREAFRGFOER 118
QY 121 ASADKTRLEAFIEAERFWLEDYALFMAKTRFDGKPNWSPPELRDRPAALAREEL 180
DB 119 ASREEKAPQFPFGTRERSWLDYTLFMAKPTHTGGLPWNRPMPRLRLEKALREASIAL 178
QY 181 ABEVALYEWIOWLFYLENGQTKAYAESKGIQIIGDMPFVAFDSSDVWNPQYFVLEADG 240
DB 179 SQEVAFHAWTOWLFPEAWKALKEEAALGIIQIGDMPFVAFDSSAEVWNPQYFVLEADG 238
QY 241 NPTVAGVPRDYFSETGOLWGNPNLYRWDVMDERDFANCIARIQSLKQCHLVRIIDHPRGF 300
DB 239 RPTVAGVPPDYFSETGOLWGNPNLYRWEVLREGEFSWIELRLKALELFLHVRIDHPRGF 298
QY 301 EAYWEVPPGRPNNAVGRWVKAPGEKLFPAARQAQSLDAPIIAEDLGVIITPEVEALRDGFGF 360
DB 299 EAYWEIPASCTAVGRWVKAPGEKLFQKIQTETGRVPIAEDLGVIITPEVEALRDGFGF 358
QY 361 PGMKILOFAPS-GEONAFPLPHNYPAHGNVYVSGTHDNDTTLGWFTAPAEARAFMAYL 419
DB 359 PGMKVLOFAPDHGMENPFLPHNYPAHGNVYVSGTHDNDTTLGWFTAPAEARAFMAYL 418
QY 420 ARYGIRCLSEVEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRGLDNWAWRYA 479
DB 419 AEWGIGFQREIEIPWALMGLKMSVARLAIYPVQDVLALGSEARMNYPGRSGNWAWEALL 478
QY 480 EGDLEPGLAAGLALAEASOR 500
DB 479 PGQLTQEHARLAWAEATGR 499

RESULT 6
Q4HAI6_9DEIO PRELIMINARY; PRT; 503 AA.
AC Q4HAI6_9DEIO
DT 13-SEP-2005 (TREMBlrel. 31, Created)
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
DE Glycoside hydrolase, family 77 (EC 2.4.1.25).
GN ORFNames=DgeODRAFT_1524;
OS Deinococcus geothermalis DSM 11300.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=319795;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=DSM 11300;
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome assembly of Deinococcus geothermalis
RT DSM 11300."
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=DSM 11300;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Deinococcus geothermalis
RT DSM 11300."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAH01000003; EAL83418.1; -; Genomic DNA.
KW Glycosyltransferase; Hydrolase; Transferase.
SQ SEQUENCE 503 AA; 57157 MW; 7982B2B2BA8AFB5D5 CRC64;

```

```

Query Match 54.3%; Score 1475; DB 2; Length 503;
Best Local Similarity 57.7%; Pred. NO. 3.2e-103;
Matches 292; Conservative 47; Mismatches 155; Indels 12; Gaps 5;

```



Qy 1 MQLQAPGILLHPTSPFRWGIGALGREAERFLDWLADAGARWVOVLPGLTSGDSDPYQ 60  
 Db 1 MPFIRSSGVLHPTSCPSRFGIDGLGEAYKFIIDFLEKSYOQYQVQVLPGLTSGVNSPYM 60  
 Qy 61 SFSAPAGNPVLDPEMLIEKMGLEQSEAP--PPYPTQVVDVGLVYOTRMPVLLRRAFAQGR 118  
 Db 61 SYSALAGNHLISPEKJLJDEGLLSDDEFAHLPNFNEKVDQVAPIKIQLLKACENFK 120  
 Qy 119 ARASQDKTRLEAFIEAERFMLEDYALFMALKTFRDGKPNWSPSPELRDREPAALARARE 178  
 Db 121 TKASPLQKQAGFCETKSYMLDDYALFMALKTQDSSSWHTWEPALAKREPDALEKQVR 180  
 Qy 179 ELAEVALYEWIOWLPYLEWGOTKAYAESKGIQIIGDMPIFVAPDSSDVWNPQVYFL-E 237  
 Db 181 QLTDEIFYKFIQYEFFRQWSELKSYANNRGIEIIGDPIYVAHDSADVWNPQVIFCLDE 240  
 Qy 238 ADGNPTVAGVPRDYFSETGOLWGNPLVYRDWMDRDNFAWCITARIQSLKQCHLVIRIDHF 297  
 Db 241 ETGEVALMAGVPPDIFSATGOLWGNPVTNWEELQKDFKWWVQREAMDYDVVIRIDHF 300  
 Qy 298 RGFEAYWEVPPGRPNNAVGRWKAFGEKLFVAVRAQLSDAPIIAEDLGVIITPEVEALRDG 357  
 Db 301 RGFEAFVTPQGEETAMNGEWTAPGESELDKQKGLPVLAEDLGVIITPEVEALRDK 360  
 Qy 358 FGFPQMKILOPAP-SCDNALFLPHYPAGHNVVYVSGTHDNDTTLGWFTAPAEAPMR 416  
 Db 361 YEFPGMKVLOPAGFDGPNPFLPNYSR--NFWVTGTHTDNDTTLGWFTANDYEKNLL 418  
 Qy 417 AYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVILGPGPARNNFPGRLGDNWAW 476  
 Db 419 LYLQ-----CISPEGIHWDLIRLASSVANQAIIVPLQDILGLNGEARNWFFSIAGKNWAW 473  
 Qy 477 RYAEGLDLEPGLAAGURALAEASQRA 501  
 Db 474 RYDFAVLTDLSRLKILTKLYGRA 498

RESULT 9  
 Q7NG57 GLOVI PRELIMINARY; PRT; 504 AA.  
 AC Q7NG57  
 DT 01-MAR-2004 (TRENBLrel. 26, Created)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE 4-alpha-glucanotransferase.  
 GN OrderedLocusNames=gll3316;  
 OS Gloeobacter violaceus.  
 OC Bacteria; Cyanobacteria; Gloeobacteria; Gloeobacterales; Gloeobacter.  
 OX NCBI\_TaxID=33072;  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RP STRAIN=PCC 7421;  
 RC MEDLINE=22977040; PubMed=14621292;  
 RX Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T., Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C., Kohara M., Matsunoto M., Matsuno A., Nakazaki N., Shingo S., Takeuchi C., Yamada M., Tabata S.;  
 RA "Complete genome structure of Gloeobacter violaceus PCC 7421, a cyanobacterium that lacks thylakoids";  
 RT DNA Res. 10:137-145(2003).  
 RL EMBL; BA000045; BAC91257.1; -; Genomic DNA.  
 DR HSSP; 087172; 1ESW.  
 DR GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.  
 DR GO; GO:0016740; P:transferase activity; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR003385; Glyco\_hydro\_77.  
 DR Pfam; PF02446; Glyco\_hydro\_77; 1.  
 DR TIGRFams; TIGR00217; malQ; 1.  
 KW Complete proteome; Transferase.  
 SQ SEQUENCE 504 AA; 57041 MW; B781C4C1AFD23BE2 CRC64;

Query Match 46.0%; Score 1248; DB 2; Length 504;  
 Best Local Similarity 49.2%; Pred. No. 5.2e-86;

Matches 252; Conservative 73; Mismatches 155; Indels 32; Gaps 10;  
 Qy 1 MQLQAPGILLHPTSPFRWGIGALGREAERFLDWLADAGARWVOVLPGLTSGDSDPYQ 60  
 Db 1 MPFIRSSGVLHPTSCPSRFGIDGLGEAYKFIIDFLEKSYOQYQVQVLPGLTSGVNSPYM 60  
 Qy 61 SFSAPAGNPVLDPEMLIEKMGLEQSEAPPP-----YPTQVVDVGLVYOTRMPVLLR 111  
 Db 61 SCFAFAGNPVLDPEMDLVAQALIDKODLADDELKTTAGWHHPKINFEAVAYKTR--LLE 118  
 Qy 112 RAFAGFRARASADKTRLEAFIEAERFMLEDYALFMALKTFRDGKPNWSPSPELRDREPA 171  
 Db 119 TACENFAAKGDS-DKLFVAYCEREGWLDYALFMALKEANDTLPQWMDQGLALREGA 177  
 Qy 172 ALAPARELAEVALYEWIOWLPYLEWGOTKAYAESKGIQIIGDMPIFVAPDSSDVWNP 231  
 Db 178 VLKQOSRELDRVYHQFSQYLSFROWSALKQVANGRSVILGDLPIYVAADSADVAHP 237  
 Qy 232 QYFYLEADGNPTVAGVPRDYFSETGOLWGNPLVYRDWMDRDNFAWCITARIQSLKQCHL 291  
 Db 238 HLFQDSEGNPTVAGVPPDIFSATGOLWGNPVTNWEEMERNGYAWINRIKRLMDWYDL 297  
 Qy 292 VRIDHFRGFAYWEVPPGRPNNAVGRWKAFGEKLFVAVRAQLSDAPIIAEDLGVIITPEV 351  
 Db 298 VRIDHFRGFAYWEVAGRRTAIDGRWIKGPGPOLDAIEKALGSLPVVABDLGLVITDEV 357  
 Qy 352 EALRDGFGFGMKILOPAPSG-EDNALFLPHYPAGHNVVYVSGTHDNDTTLGWFTAP-- 408  
 Db 358 ETLRDYGLGFMKVLQVAPDGMANNIHLPHNHR--SSIVVTGTHTDNDTTLGWYHNSPRW 415  
 Qy 409 --EAERAFMAYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVILGPGPARNNF 466  
 Db 416 VHEATERYL--YLN-----TGWEIHWALVRCALASVCNLSILPMDQVILGSGEARNNT 466  
 Qy 467 PGRGLDNWAWRYAEGD-LEPGL---AAGLRAL 494  
 Db 467 PGVAGNWDNRVGTTFEAVPEWMRERLADLTAL 498

RESULT 10  
 Q74DY3 GEOSL PRELIMINARY; PRT; 493 AA.  
 AC Q74DY3  
 DT 05-JUL-2004 (TRENBLrel. 27, Created)  
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)  
 DE 4-alpha-glucanotransferase (EC 2.4.1.25).  
 GN Name=malQ; OrderedLocusNames=GSU1182;  
 OS Geobacter sulfurreducens.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;  
 OC Geobacteraceae; Geobacter.  
 OX NCBI\_TaxID=35554;  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=PCC / ATCC 51573;  
 RC PubMed=14671304; DOI=10.1126/science.1088727;  
 RX Meche B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C., Heidelberg J.F., Wu D., Ward N.L., Beanan M.J., Dodson R.J., Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S., Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J., Davidson T.M., Zafar N., White O., Tran B., Romero C., Forberger H.A., Weidman J.F., Kouri H.M., Feldblyum T.V., Utterback T.R., Van Aken S.E., Lovley D.R., Fraser C.M.;  
 RA "Genome of Geobacter sulfurreducens: metal reduction in subsurface environments";  
 RT Science 302:1967-1969(2003).  
 RL EMBL; AB017180; AAR34558.1; -; Genomic DNA.  
 DR TIGR; GSU1182; -  
 DR GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.  
 DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR003385; Glyco\_hydro\_77.  
 DR Pfam; PF02446; Glyco\_hydro\_77; 1.





OrderedLocusNames=tl:0708;  
OS Synechococcus elongatus (Thermosynechococcus elongatus).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=32046;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=BP-1;  
RX MEDLINE=22225144; PubMed=12240834;  
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,  
Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakaraki N.,  
Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the thermophilic cyanobacterium  
Thermosynechococcus elongatus BP-1";  
RL DNA Res. 9:123-130(2002).  
DR EMBL; BA000039; BAC08259.1; -; Genomic DNA.  
DR HSP; 087172; 1ESW.  
DR GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR003385; Glyco\_hydro\_77.  
DR Pfam; PF02446; Glyco\_hydro\_77; 1.  
DR TIGRFAMs; TIGR00217; malQ; 1.  
KW Complete proteome; Transferase.  
SQ SEQUENCE 518 AA; 58347 MW; 3DA8B8A5DF2474B1 CRC64;  
Query Match 44.8%; Score 1217; DB 2; Length 518;  
Best Local Similarity 49.8%; Pred. No. 1.2e-83;  
Matches 249; Conservative 61; Mismatches 174; Indels 16; Gaps 6;  
QY 1 MOLQAFGILLHPTSPGRWGIGALGREARFLDLADAGARWQVPLGTSYDSDPYQ 60  
DB 1 MHFPRCGLLHPTSPGGHIGELGTAAEFLEFLVASQVQVQLGPTGSGNSPYM 60  
QY 61 SPSAFAGNPLYVDPEMLIEKWLQSEAPPPYPT- - - - -QRVDYGVLYQTRWPLLRPAFAG 116  
DB 61 CYSAMAGNPLILSLEEVAKAGWL- - - - -TEADLGQITLENRDRVDFDAVISQKLLRLAAQR 118  
QY 117 FRASAQDKTRLEAFIEARFWLEDYALFMALKTRFDGKPNWSPDELDRPAALARA 176  
DB 119 FQSQATPGDQWAFRDFQALAHYLPNVALFMALKHGOGFWTEWPAFLDRPETAALAI 178  
QY 177 RESELAEEVALYEQWLFWLEWQGTQKAYAESKGIQIGDMPFIYVAFDSSDVWNPQYFYL 236  
DB 179 QVVLKDRIFEEYEQQLFYQWHLKAEANQGIQIIGDPIYVAHDSVDWVAPQPFEL 238  
QY 237 EAD- GNPVVAGVPRDYFSTGQWGNPLRWDMERDNFANCIARIQSLKQCHLVRI 295  
DB 239 NPETGAALMAGVPPDYFSATGQWGNPIYVWALADGYSWIERFRALLAYVDIIRVD 298  
QY 296 HFRGFENYVPPGRPNVAGRWKAPGEKLFPAVRAQLSDAPIAEDLGVTITPEVALR 355  
DB 299 HFRGFQAYWQVPEKTAQVNGQPGGAFFALQALQRLPILAEGLDITPDVALR 358  
QY 356 DGFQFQMKILOPAF- SGEDNAFLPHNYPAHGNVNVYSGTHDNDITLWGFRTAPEAF 414  
DB 359 DQFQFQMKILOPAFQGGSDNPLFPN- - - - -QERNVNVYSGTHDNDITLWGFRTAPEAF 416  
QY 415 MRAYLARYGIRCLSEYEVACALIELAFKSPAKLAIVPLQVGLGPGPARNFQRLGDNW 474  
DB 417 FIDYLG- - - - -YTPSPHVALIRMALGTVAQAIIPVQDLGLDLSHARNFPDQGNW 470  
QY 475 AWRYAEGDLEPGLAAGLRAL 494  
DB 471 AWRLTPGQLTPELAHLRL 490

RESULT 14  
MALQ STRPN  
ID MALQ STRPN STANDARD; PRT; 505 AA.  
AC POA300; P29851;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)

NUCLEOTIDE SEQUENCE.  
RC STRAIN=2CP-C;  
RG US DOE Joint Genome Institute (JGI-P&F);  
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,  
Hammon N., Israni S., Pitluck S., Richardson P.;  
RT "Sequencing of the draft genome assembly of Anaeromyxobacter  
dehalogenans 2CP-C";  
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=2CP-C;  
RX US DOE Joint Genome Institute (JGI-ORNL);  
RA Larimer F., Land M.;  
RT "Annotation of the draft genome assembly of Anaeromyxobacter  
dehalogenans 2CP-C";  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
preliminary data.  
DR EMBL; AHD0100014; EAL79233.1; -; Genomic DNA.  
DR Glycosyltransferase; Hydrolase; Transferase.  
KW SEQUENCE 518 AA; 56400 MW; A6CF058B1AB9C586 CRC64;  
Query Match 45.0%; Score 1221; DB 2; Length 518;  
Best Local Similarity 49.8%; Pred. No. 6e-84;  
Matches 255; Conservative 53; Mismatches 180; Indels 24; Gaps 7;  
QY 5 RAFGILLHPTSPGRWGIGALGREARFLDLADAGARWQVPLGTSYDSDPYQFSA 64  
DB 4 RRSGLLHPTSLPGHAGDGLAAHRRPAGWLAAGQLWVPLGPTGFGDSYQALSS 63  
QY 65 FAGNPLYVDPEMLIEKWLQSE- - - - -APPPYPTQYDVGWLYQTRWPLLRPAFAGFARA 121  
DB 64 HAGNPLVLSLEVNQGLDADLSGAPAGEP-GRADLHAALWKRRERLARAARAFRAGA 122  
QY 122 SAQDKTRLEAFIEARFWLEDYALFMALKTRFDGKPNWSPDELDRPAALARBELA 181  
DB 123 DGERAAELEDPRAREAGWLEDWALFAALKAAHGGRPWTAPPLARRERAALESARARFA 182  
QY 182 EEVALYEQWLFWLEWQGTQKAYAESKGIQIGDMPFIYVAFDSSDVWNPQYFYLEADGN 241  
DB 183 HEVEAEVAFQAFHQAARALRCALGIELMGDLPIYVAHDSAEVWAPPELFRDLAAGE 242  
QY 242 PTVVAGVPRDYFSTGQWGNPLRWDMERDNFANCIARIQSLKQCHLVRIHFRGFE 301  
DB 243 PAAVAGVPPDYFSATGQWGNPLRWDMERDNFANCIARIQSLKQCHLVRIHFRGFE 302  
QY 302 AYWEVPPGRPNVAGRWKAPGEKLFPAVRAQLSDAPIAEDLGVTITPEVALRGGFPP 361  
DB 303 AYWEIPAGAPTAERGRWVPGGARLFEALLRALGPLPFAENGLVITPEVALRRRFLP 362  
QY 362 GSKTLOFAFGSDNA- - - - -FLPHNYPAHGNVNVYSGTHDNDITLWGF- - - - -RTAPEA- 410  
DB 363 GMAILQFAGDQPAFTQPHNYAP- - - - -DLVAYTGHNDITVAGWGGAGGSDVTAEEVA 420  
QY 411 -ERAFMAYLARYGIRCLSEYEVACALIELAFKSPAKLAIVPLQVGLGPGPARNFQRLGDNW 469  
DB 421 REKFALEYLVGDG- - - - -RDVPGAMIRAVLASVADTVFPLQDALGLGSEARNMTAT 474  
QY 470 LGDNWARYAEGDLEPGLAAGLRALAEASQA 501  
DB 475 LGGNRWRVREAEALDAALAAARLRRLAAVYGRA 506

RESULT 13  
Q8DKZ1 SYNEL  
ID Q8DKZ1 SYNEL PRELIMINARY; PRT; 518 AA.  
AC Q8DKZ1  
DT 01-MAR-2003 (TReMBLrel. 23, Created)  
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
DE 4-alpha-glucanotransferase.



Query Match	44.0%;	Score 1194;	DB 1;	Length 505;
Best Local Similarity	48.1%;	Pred. No. 6.5e-82;		
Matches 242;	Conservative 77;	Mismatches 166;	Indels 16;	Gaps 9;
QY	4	QRAFGLIHLPTSPPGRWGICALGREAEFLDLWLADAGARWQVLPGLGPTS YGDSPYQSFS	63	
Db	3	KRQSGVLHHSILPGAYGIGSFGQSAVDVDFLVRTKQRYWQILPLGATSYGDSPYQSFS	62	
QY	64	AFAGNPVLVDPEMLIEKWLQSEAPP---PYPTQRVDYGLWLYQYTRPFLRRAPAGFRAR	120	
Db	63	AFAGNTHFDLIDLVEQGLLEASDLGVDGSDASEVDYAKIYYARRPLLEKAVKRFEV	122	
QY	121	ASADKTRLEAFIEAERFWLEDYALFWALKTRFDGKPNW-SPELRDRREPAALARAREE	179	
Db	123	GDVKD---FEKPAQDNOSWLELFAEYMAKEYFDNLAWTEWPDADARAKASALESTREQ	179	
QY	180	LAEEVALYEWIQWLFYLEWQGTAKAYAESKGIQIIGDMPIFVAFDSSDVWANPOYFYLEAD	239	
Db	180	LADKLVYHRVTQYFFQOMLKLKAYANDNHIEIVGDMPIYVAEDSSDMWANPHLFKTDVN	239	
QY	240	GNPTVWAGVPRDYSETGQLWGNPLRYRWDVNERDNFAWCIARIQSLKQCHLVRIHDFRG	299	
Db	240	GKATCIAGCPDPEFSVTQQLWGNPTDYWEAMDKGKWWIERLRESFKIYDIVRIDHFRG	299	
QY	300	FEAYNEVPFGPRNAVEGHWKAPGKLPAAVRAQLSDAPIIAEDLGVTTPREVEALRDGFG	359	
Db	300	FESYWEIPAGSDTAAPGEWVGPGYKLPAAVKEELGELNIIAEDLGFMIDEVIELRBTG	359	
QY	360	PPGKILQAFSGEDNAP-LPHNTYPAHGNVYVYSGTHDNTTLGNFRTAPEAEAFMRAY	418	
Db	360	PPGKILQAFAPNPEDESIDSPHLAPA--NSVMYTGTHDNTVLGWYRN--EIDDA-TREY	414	
QY	419	LARYGIRCLSEYE-VAGALIELAFKSPAKLAIPLQDVGLGPEARNFPGRLGDNWNR	477	
Db	415	MARYTNR--KEYETVWHLMTVFSSVSFMAIATWQDLLEDEAARMNFPSTLGNWSR	472	
QY	478	YAEGLERPLAAGLALAEASQR	500	
Db	473	MTEDQLTFAVEEGLDLTTIYR	495	

Search completed: January 10, 2006, 14:12:38  
Job time : 167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2006, 14:06:43 ; Search time 47 seconds  
(without alignments)  
881.288 Million cell updates/sec

Title: US-10-618-976-2

Perfect score: 2714

Sequence: 1 MQLQAFGILLHPTSPFGRW.....DLRGLAAGLALAEASORA 501

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82575679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*
- 1: /cgn2\_6/ptodata/1/iaa/5 COMB.pap.\*
  - 2: /cgn2\_6/ptodata/1/iaa/6 COMB.pap.\*
  - 3: /cgn2\_6/ptodata/1/iaa/H COMB.pap.\*
  - 4: /cgn2\_6/ptodata/1/iaa/PCUS COMB.pap.\*
  - 5: /cgn2\_6/ptodata/1/iaa/RE COMB.pap.\*
  - 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2714	100.0	501	2	US-09-687-360-2
2	1196	44.1	505	2	US-09-583-110-4228
3	1076	39.6	458	2	US-09-107-433-2939
4	777.5	28.6	298	1	US-08-838-543-5
5	694	25.6	291	1	US-08-838-543-6
6	606.5	22.3	470	2	US-09-902-540-9896
7	451	16.6	526	2	US-09-198-452A-347
8	451	16.6	536	2	US-09-438-185A-328
9	299	11.0	681	2	US-09-252-991A-24629
10	278	10.2	743	2	US-09-489-039A-11073
11	273.5	10.1	496	1	US-08-838-543-2
12	197	7.3	323	1	US-08-838-543-4
13	187	6.9	330	1	US-08-838-543-3
14	123	4.5	1332	2	US-09-252-991A-25772
15	105.5	3.9	508	2	US-09-583-110-5034
16	105.5	3.9	594	2	US-09-107-433-4327
17	104	3.8	436	2	US-09-252-991A-17273
18	101.5	3.7	926	2	US-08-855-910-2
19	101	3.7	721	2	US-09-583-110-5179
20	101	3.7	750	2	US-09-107-433-4321
21	101	3.7	1698	2	US-09-902-540-12813
22	100.5	3.7	468	2	US-09-832-129-59
23	100	3.7	569	2	US-08-961-083-154
24	100	3.7	569	2	US-09-536-784-154
25	100	3.7	569	2	US-09-765-271-154
26	100	3.7	569	2	US-09-765-272A-154
27	100	3.7	658	2	US-09-769-787-17

28	100	3.7	677	2	US-09-769-787-155	Sequence 155, Appl
29	98	3.6	752	2	US-09-252-991A-30127	Sequence 30127, A
30	97.5	3.6	511	2	US-09-266-965-127	Sequence 127, Appl
31	97.5	3.6	1047	2	US-09-252-991A-29125	Sequence 29125, A
32	96.5	3.6	432	2	US-09-832-129-41	Sequence 41, Appl
33	96.5	3.6	432	2	US-09-832-129-54	Sequence 54, Appl
34	96.5	3.6	438	2	US-09-167-299-5	Sequence 5, Appl
35	96.5	3.6	4545	1	US-08-804-227C-14	Sequence 14, Appl
36	96.5	3.6	4550	1	US-08-804-227C-8	Sequence 8, Appl
37	96.5	3.6	4550	1	US-08-804-198-2	Sequence 2, Appl
38	96	3.5	823	2	US-09-252-991A-24768	Sequence 24768, A
39	95.5	3.5	301	2	US-09-148-545-232	Sequence 232, Appl
40	95.5	3.5	301	2	US-09-621-011-232	Sequence 232, Appl
41	95.5	3.5	639	2	US-09-252-991A-25089	Sequence 25089, A
42	95	3.5	591	2	US-08-961-083-74	Sequence 74, Appl
43	95	3.5	591	2	US-09-536-784-74	Sequence 74, Appl
44	95	3.5	591	2	US-09-765-271-74	Sequence 74, Appl
45	95	3.5	591	2	US-09-765-272A-74	Sequence 74, Appl

ALIGNMENTS

RESULT 1  
US-09-687-360-2  
; Sequence 2, Application US/09687360  
; Patent No. 6617143  
; GENERAL INFORMATION:  
; APPLICANT: Fukuyama, Shiro  
; TITLE OF INVENTION: Polypeptides Having Glucanotransferase  
; FILE OF INVENTION: Activity And Nucleic Acids Encoding Same  
; FILE REFERENCE: 6012.200-US  
; CURRENT APPLICATION NUMBER: US/09/687,360  
; CURRENT FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: PA 1999 01501  
; PRIOR FILING DATE: 1999-10-20  
; PRIOR APPLICATION NUMBER: PA 1999 01641  
; PRIOR FILING DATE: 1999-11-15  
; PRIOR APPLICATION NUMBER: 60/160,903  
; PRIOR FILING DATE: 1999-10-22  
; PRIOR APPLICATION NUMBER: 60/166,539  
; PRIOR FILING DATE: 1999-11-18  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 501  
; TYPE: PRT  
; ORGANISM: Thermus rubens  
US-09-687-360-2

Query Match	100.0%;	Score 2714;	DB 2;	Length 501;
Best Local Similarity	100.0%;	Pred. No. 6.5e-252;		
Matches 501;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	MQLQAFGILLHPTSPFGRWGIGALGREAERFDLWDLADAGARWQVPLGPTSYGDSFYQ	60	
DB	1	MQLQAFGILLHPTSPFGRWGIGALGREAERFDLWDLADAGARWQVPLGPTSYGDSFYQ	60	
QY	61	SFSAFAGNPYLDVPEMLIEKGMLEQSEAPPPYQRTQVDYGNLYQTRWPLLRAPAGFRAR	120	
DB	61	SFSAFAGNPYLDVPEMLIEKGMLEQSEAPPPYQRTQVDYGNLYQTRWPLLRAPAGFRAR	120	
QY	121	ASQDKTRLEAFTEARERFWDYALFMALKTRFDGKPMWNSPELRDPAALARAEEEL	180	
DB	121	ASQDKTRLEAFTEARERFWDYALFMALKTRFDGKPMWNSPELRDPAALARAEEEL	180	
QY	181	AEBAVALYEWIOMLYLEWGTQKAYAESKGIQIIGDMPIFVAFDSSDVWNPQYFYLEADG	240	
DB	181	AEBAVALYEWIOMLYLEWGTQKAYAESKGIQIIGDMPIFVAFDSSDVWNPQYFYLEADG	240	
QY	241	NPTVAVGPRDYPSETGOLMGNPNLYRWDMERNFNAMCIARIQSLKQCHLVRIIDHPRGF	300	
DB	241	NPTVAVGPRDYPSETGOLMGNPNLYRWDMERNFNAMCIARIQSLKQCHLVRIIDHPRGF	300	

QY 301 EAYVEVPGRNVEGRWVKA PGSKLFAAVRAQLSDAPIIAEDLGVTTPVEALRDGFGF 360  
 DB 301 EAYVEVPGRNVEGRWVKA PGSKLFAAVRAQLSDAPIIAEDLGVTTPVEALRDGFGF 360  
 QY 361 PGMKILQFAPSGEDNAPLPHNYPAHGNVVVYSGTHDNDTTLGWERTAPAEARAFMAYLA 420  
 DB 361 PGMKILQFAPSGEDNAPLPHNYPAHGNVVVYSGTHDNDTTLGWERTAPAEARAFMAYLA 420  
 QY 421 RYIGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFFGRGLGDNWAWYAE 480  
 DB 421 RYIGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFFGRGLGDNWAWYAE 480  
 QY 481 GDLEPGLAAGLALAEASORA 501  
 DB 481 GDLEPGLAAGLALAEASORA 501

RESULT 2  
 US-09-583-110-4228  
 ; Sequence 4228, Application US/09583110  
 ; Patent No. 6699703  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al.  
 ; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
 ; FILE REFERENCE: PATH00-07A  
 ; CURRENT APPLICATION NUMBER: US/09/583,110  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR FILING DATE: 1998-06-30  
 ; PRIOR FILING DATE: 1998-05-12  
 ; PRIOR FILING DATE: 1997-07-02  
 ; NUMBER OF SEQ ID NOS: 5322  
 ; SEQ ID NO 4228  
 ; LENGTH: 505  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus pneumoniae  
 US-09-583-110-4228

Query Match 44.1%; Score 1196; DB 2; Length 505;  
 Best Local Similarity 48.1%; Pred. No. 4.1e-106; Indels 16; Gaps 9;  
 Matches 242; Conservative 77; Mismatches 168

QY 4 ORAFGILLHPTSFGRMGIGALGREAEFLDWLADAGARWQVLPLGPTSYGDSFYQSFS 63  
 DB 3 KRQSGVLMHISLPGAYGIGSGOSAYDFVFLVTRKQRYWQILPLGATSYGDSFYQSFS 62  
 QY 64 AFAGNPYLVDPEMLIEKGWLSQSEAPP---PYPTQVDYGLWYOTRPLLRAPAGFRAR 120  
 DB 63 AFAGNTHFDLIDLVVEQGLLEASDLGVDGSDASEVDYAKIYYARRPLLEKAVKRFEEV 122  
 QY 121 ASADKTRLEAFIEARERWEDYALFMALKTRFPCKPNW-SPELDRPEALAREEE 179  
 DB 123 GDVKD---FEKFAQDNQSMLELFAEYMAIKEHFDNLAWTEWPDADARKASALESTREQ 179  
 QY 180 LAEEVALYEMTQWLFLYEWGQTKAYAESKGIQIGDMEIFVAFDSSDVWNPQVYLEAD 239  
 DB 180 LADKLIVHRVTQYFFQWKLKAYANDNHIIEVGDMPYIYAEDESSDMWNPFLFKTDVN 239  
 QY 240 GNPTVAGVPDRYSETQOLWGNPLRYWVDWNERNFANCIARISQSLKQCHLVIRIDHPRG 299  
 DB 240 GKATCIAGCPDPEFSVTQOLWGNPIYWEAMDKGKWKWIERLRESFKIYDIVRIDHPRG 299  
 QY 300 FEATWEVPFGPNVAGEWVKA PGSKLFAAVRAQLSDAPIIAEDLGVTTPVEALRDGFG 359  
 DB 300 FESYWEIPAGSDTAAPGWEVGPYKLFPAVKEELGELNLIIAEDLGFWTDEIELERTG 359  
 QY 360 FPGMKILQFAPSGEDNAPLPHNYPAHGNVVVYSGTHDNDTTLGWERTAPAEARAFMAY 418  
 DB 360 FPGMKILQFAPSGEDNAPLPHNYPAHGNVVVYSGTHDNDTTLGWERTAPAEARAFMAY 418

QY 419 LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFFGRGLGDNWAWR 477  
 DB 415 MARYTNR---KEYETVHAMLRVTVFSSVFWAIATMQLLEDEARARNMFFSTLGGNWSR 472  
 QY 478 YARGDLEPGLAAGLALAEASOR 500  
 DB 473 MTEDQLTPAVEGLDLDTTIYRR 495

RESULT 3  
 US-09-107-433-2939  
 ; Sequence 2939, Application US/09107433  
 ; Patent No. 6800744  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
 ; THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 5206  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 ; STREET: 100 Beaver Street  
 ; CITY: Waltham  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02354  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-ROM ISO9660  
 ; COMPUTER: <Unknown>  
 ; OPERATING SYSTEM: <Unknown>  
 ; SOFTWARE: <Unknown>  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/107,433  
 ; FILING DATE: 30-Jun-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/ 085131  
 ; FILING DATE: May 12, 1998  
 ; APPLICATION NUMBER: 60/051553  
 ; FILING DATE: July 2, 1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ariniello, Pamela Deneke  
 ; REGISTRATION NUMBER: 40,489  
 ; REFERENCE/DOCKET NUMBER: GTC-011  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (781)893-5007  
 ; TELEFAX: (781)893-8277  
 ; INFORMATION FOR SEQ ID NO: 2939:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 458 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: YES  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Streptococcus pneumoniae  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (B) LOCATION 1...458  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 2939:  
 US-09-107-433-2939

Query Match 39.6%; Score 1076; DB 2; Length 458;  
 Best Local Similarity 48.4%; Pred. No. 1.2e-94;  
 Matches 221; Conservative 66; Mismatches 154; Indels 16; Gaps 9;

QY 50 GPTSYGDSFYQSFSAPAGNPYLVDPEMLIEKGWLSQSEAPP---PYPTQVDYGLWYOTR 106  
 DB 2 GATSYGDSFYQSFSAPAGNPYLVDPEMLIEKGWLSQSEAPP---PYPTQVDYGLWYOTR 61  
 QY 107 WPLLRAPAGFRARASADKTRLEAFIEARERWEDYALFMALKTRFPCKPNW-SPEL 165  
 DB 62 RPLLEKAVKRFEEVGDVKD---FEKFAQDNQSMLELFAEYMAIKEHFDNLAWTEWPDADA 118







QY 20 WGIGALGREAEKFLDWLADAGRW-----WQVFLGFLSISGDSFPIQSFSFAGNFILVD 73

QY 20 WGIGALGREAEERFLDWLADAGRW-----WQVFLGFLISIGUSFINSFSAFAGNFLLVD 75

[illegible]

RESULT 12  
US-08-838-543-4  
; Sequence 4, Application US/08838543  
; Patent No. 5994623  
; GENERAL INFORMATION:  
; APPLICANT: KREBBERS, ENNO  
; APPLICANT: BROGLIE, KAREN E.  
; TITLE OF INVENTION: CORN 4-(-GLUCANOTRANSFERASE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.50 INCH DISKETTE  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (VER. 7.0)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/838,543  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MAJARIAN, WILLIAM R.  
; REGISTRATION NUMBER: P-41,173  
; REFERENCE/DOCKET NUMBER: BB-1101  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-992-4926  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 4:

```

206 WGIGDFG-DLRAMLPEIARRGGSIGLNPihalYPANPES--ASPYSPSSRWLNVIYID 262
74 PEMLIKGNLEQSEAPPYP-----TORVDYCWLYQTRWPLLRRAFAFRARAS 122
263 VNAVEDPQRSEBAQAWQSPATQALQARETDDVDYAVTTLKMTALRMAWKQFSRRED 322
123 AQDKTRLEAFTEAS--RFWLEDYALFMALTRFDGKPN--EWSPELRDREPAALARAR 177
323 EQ-WTAFREFVLRGESLYWQAADFALHAWQVDPLRWGWPAPKAFQDIDSPEVKAF 381
178 EEALAEVYALYEWIOWLYFLEWQYKVAESK-----GIQIGDMPIFVAFPDSSDVWANPQ 232
382 VEHEDDVSFYLWLQW--LAWSOFAACWETSQDQMPIGLYRDLAVGVAEQSGETWCDORE 438
233 YFYLEADGNPTVAGVRDYESGTQMGWGNPLYRWDMVNERDNFAWCIAIRIQSLKQCHLV 292
439 LYCLUKAS-----VCAPDIIIGPIQWGNGLPMPDPIIITAARAYEPFIDLLRANQWCGAL 492
293 RIDHFRGFEAYWYEPFRPNAVEGRWVKAPGEKLFU--AVRAQLSDAFIIAEDLGVIITPE 350
493 RIDHVMSVLRWLWPIYGE-TADHGAYVQVPVDDLSSLALSALESQRHRCMVIGEDLTGVE 551
351 VEALRDGFGPGMKILQAFSGEDNAFIPHYPAHGNVVVYSGTHDNDTTLGWFTAPEA 410
552 IVSKLRXSGVYSYKVLVPESDAEKTFRAPALYPEQSMAY--ATTHDLTPLRGYWSGGOLT 609
411 ERAFWRAY-----LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQD-- 454
610 LGKALGLYDPDEVLRGLYQDRELAKQGH--LDLALHKYGLCPKRGHKAHSLMSMTGILNRG 667
455 -----VLGLGPE-----ARMNFPGRLGDNWAWR 477
668 MORYIADNSALIGIOPDWLEMAFVNIPTGTSFTEYPNWR 707

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 323 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-838-543-4

```

[illegible]

RESULT 13  
US-08-838-543-3  
Sequence 3, Application US/08838543  
Patent No. 5994623  
GENERAL INFORMATION:  
APPLICANT: KREBBERS, ENNO  
APPLICANT: BROGLIE, KAREN E.  
TITLE OF INVENTION: CORN 4-(-GLUCANOTRANSFERASE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.50 INCH DISKETTE  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WINDOWS 95  
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (VER. 7.0)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/838,543  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: MAJARIAN, WILLIAM R.  
REGISTRATION NUMBER: P-41,173  
REFERENCE/DOCKET NUMBER: BB-1101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-992-4926  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 330 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-838-543-3

Query Match 6.9%; Score 187; DB 1; Length 330;  
Best Local Similarity 23.1%; Pred. No. 1.8e-09;  
Matches 70; Conservative 43; Mismatches 144; Indels 46; Gaps 8;

[illegible]

RESULT 14  
US-09-252-991A-25772  
US-09-252-991A-25772, Application US/09252991A  
; Sequence 25772, Patent No. 6551795  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25772  
; LENGTH: 1332  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25772

Query Match	4.5%;	Score 123;	DB 2;	Length 1332;
Best Local Similarity	19.6%;	Pred. No. 0.02;		
Matches 103;	Conservative 52;	Mismatches 155;	Indels 216;	Gaps 25

Qy	17	PCRNGICALGREAERF-----LDWLADAGARWQVPLPGPTSYGDSPPQSFQSFSAFAGN	68
Db	54	PCRGCAGAMS-GATLFMLVLLVLLADLLGALGWWR-----TRGTEIRSF-----	99
Qy	69	PYLVDPMLETKWGLEQSEAPPPYPTQRVDYGLYQ-----TRW-----PLLRRAF	114
Db	100	YAAVRQWEREQWQGRYESP-----WLLMGNETEGEQLCATWRLPVARPAW	147
Qy	115	AGFRARASADKTRL-----EAFIEARF-----WLEDYALFMAKTR--FDGKPMWNSP	163
Db	148	FG-RWNSDGEGAILVPESVFLPDEGLRQSGAWRLRLFLRGRALDGVVWNIPLA	206
Qy	164	ELDRREPAA---LAPARE--ELAEVALYEWI QWLFYLEWGQTKYAESKGIQIGMPI	218
Db	207	RLQDGEQAANLGLAARRRYVELTQRLGL-----SLVP	238
Qy	219	FVAIFDSSDVWANPQYFYFLBADGNPTVAVGPRDYFSETGLQWGNPLYRWDVMERNDFAWC	278
Db	239	YVYVITGLE-----DLPFGQELLAAUPEE-----	261
Qy	279	IARIROSLKOCHLVRIIDHFRGFAYWEVPPGRPNVAVEGRWVTAPEGLKFAAVRAQLSDAP	338

Db 262 -ARER-----ALGWSPPFAAEAAWQCEQALEEITATLTESIVELG 303  
Qy 339 I-----IAEDLGVIPEVEALRDGFGFCMKILOFAPFSGEDNAFLPHNYPAGHNVVYSGT 394  
Db 304 TLRGQVDNELYCLPRLESUGSL-----QALLEPVFOGN-----ARGEAPRFRGL 349  
Qy 395 HDNDTTILGWFRTPAPEABAFRAYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLOD 454  
Db 350 Y-----LSGSEAGAAGAAD-----DVLPAVD 369  
Qy 455 VLGPGPARMNFPGRLGDNWAWRYAEGDLEPGLAAGLRALABASQR 500  
Db 370 ----APRRRSFASQL---WARRIL---AEEGLAQAVPRILQLROR 405

RESULT 15  
US-09-583-110-5034  
; Sequence 5034, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583.110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 5034  
; LENGTH: 508  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-5034

Query Match 3.9%; Score 105.5; DB 2; Length 508;  
Best Local Similarity 21.8%; Pred. No. 0.23;  
Matches 73; Conservative 51; Mismatches 106; Indels 105; Gaps 22;  
Qy 51 PTSYGDSPYQSFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRYDYGLYQTRWPLL 110  
Db 95 PLFTSDGPWRA-TLKAGT--LIEEDLFTGNF--GSKAPYNF-SORQEFFDEHGKXWPLM 148  
Qy 111 RRAFAGFRARASAOQKTRLEAFIEARFWLEDYALFMALKTRFDGKPH-NEWSPELRDRE 169  
Db 149 -----CMFEW-----DG--WFNRWKEPIITRD 168  
Qy 170 PAALARAREELAEVALYEWQWLYLEWQTKAYAESKGIQIG--DMPIFVAFDSSDV 227  
Db 169 PRELADAVREVLEQGSIN-----LYMPHGGTN-FGFMNGCSARTGLDPQVTSYDYDAL 221  
Qy 228 WANPQFYLEADGNPT---VVAGVPRDYFSETGOLWGNPLNRYMDVMERNDFANCARIR 283  
Db 222 -----LDEKRNFTAKYLVKQWATHFSEYPOL--EPLYK-ESMELDAIP-LVEKVS 269  
Qy 284 -----QSLKQCHLVRIHFRGF-----EAYWE-----VPFGRPN-----VEG 316  
Db 270 LFETLDSLSPVESLYPQMBELGQSYGYLLYRTETNWDABEERLIIDGRDRAQLYVDG 329  
Qy 317 RWVKAP-----GEKLP-AAVRAQLSDAPIIAEDLG 345  
Db 330 QWVKTYQYTEIGEDIFYQKKKGLSLDLITENMG 364

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: January 10, 2006, 14:12:49 ; Search time 115 Seconds  
(without alignments)  
1820.283 Million cell updates/sec

Title: US-10-618-976-2

Perfect score: 2714

Sequence: 1 MOLQRAFGILLHPTSPGRW.....DLRPLAAGLALAEASQRA 501

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA Main:

- 1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /cgn2\_6/prodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2714	100.0	501	5	US-10-618-976-2
2	1194	44.0	505	5	US-10-472-928-4402
3	1076	39.6	458	5	US-10-617-320-2939
4	989	36.4	567	4	US-10-424-599-170978
5	894.5	33.0	486	4	US-10-425-114-57203
6	777.5	28.6	298	4	US-10-336-597-5
7	767.5	28.3	486	4	US-10-425-115-325328
8	694	25.6	291	4	US-10-336-597-6
9	477	17.6	243	4	US-10-437-963-138565
10	451	16.6	526	3	US-09-841-132-518
11	451	16.6	526	4	US-10-289-762-347
12	451	16.6	526	5	US-10-872-155-518
13	430.5	15.9	527	3	US-09-841-132-518
14	430.5	15.9	527	5	US-10-872-155-518
15	412.5	15.2	179	4	US-10-767-701-51019
16	383	14.1	583	4	US-10-081-872-188
17	383	14.1	683	4	US-10-385-303-188
18	373	13.7	740	4	US-10-156-761-12925
19	342.5	12.6	187	4	US-10-424-599-170977
20	338	12.5	946	4	US-10-437-963-112596
21	333	12.3	183	4	US-10-336-597-8
22	323.5	11.9	204	4	US-10-425-114-53388
23	313.5	11.6	684	4	US-10-389-647-478
24	273.5	10.1	496	4	US-10-336-597-2
25	272	10.0	153	4	US-10-091-007-116
26	270	9.9	336	4	US-10-425-115-357932
27	238	8.8	204	4	US-10-767-701-40716

28	237.5	8.8	496	4	US-10-424-599-155954	Sequence 155954,
29	207	7.6	98	4	US-10-425-115-234981	Sequence 234981,
30	207	7.6	102	4	US-10-336-597-16	Sequence 16, Appl
31	207	7.6	700	4	US-10-781-014-412	Sequence 412, App
32	207	7.6	706	3	US-09-738-626-6019	Sequence 6019, Ap
33	207	7.6	706	4	US-10-781-014-410	Sequence 410, App
34	197	7.3	323	4	US-10-336-597-4	Sequence 4, Appli
35	187	6.9	330	4	US-10-336-597-3	Sequence 3, Appli
36	146	5.4	185	4	US-10-424-599-223106	Sequence 223106,
37	136	5.0	123	4	US-10-425-115-357934	Sequence 357934,
38	112	4.1	678	3	US-09-738-626-6019	Sequence 6025, Ap
39	111	4.1	595	5	US-10-472-928-4832	Sequence 4832, Ap
40	109	4.0	402	4	US-10-437-963-191559	Sequence 191559,
41	107.5	4.0	652	4	US-10-369-493-18004	Sequence 18004, A
42	105.5	3.9	284	4	US-10-425-115-367937	Sequence 367937,
43	105.5	3.9	594	5	US-10-617-320-4327	Sequence 4327, Ap
44	105	3.9	86	4	US-10-767-701-39064	Sequence 39064, A
45	105	3.9	185	4	US-10-767-701-55139	Sequence 55139, A

## ALIGNMENTS

### RESULT 1

US-10-618-976-2  
; Sequence 2, Application US/10618976  
; Publication No. US20050074769A1  
; GENERAL INFORMATION:  
; APPLICANT: Fukuyama, Shiro  
; TITLE OF INVENTION: Polypeptides Having Glucanotransferase  
; FILE REFERENCE: 6012 200-US  
; CURRENT APPLICATION NUMBER: US/10/618,976  
; CURRENT FILING DATE: 2003-07-14  
; PRIOR APPLICATION NUMBER: US/09/687,360  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: PA 1999 01501  
; PRIOR FILING DATE: 1999-10-20  
; PRIOR APPLICATION NUMBER: PA 1999 01641  
; PRIOR FILING DATE: 1999-11-15  
; PRIOR APPLICATION NUMBER: 60/160,903  
; PRIOR FILING DATE: 1999-10-22  
; PRIOR APPLICATION NUMBER: 60/166,539  
; PRIOR FILING DATE: 1999-11-18  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 501  
; TYPE: PRT  
; ORGANISM: Thermus rubens  
US-10-618-976-2

Query Match 100.0%; Score 2714; DB 5; Length 501;  
Best Local Similarity 100.0%; Pred. No. 1e-245;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MOLQRAFGILLHPTSPGRWIGALGREARFLDMLADAGARWQVLPGLPSTSYGDSPIQ 60		
DB	1	MOLQRAFGILLHPTSPGRWIGALGREARFLDMLADAGARWQVLPGLPSTSYGDSPIQ 60		
QY	61	SFSAPAGNPYLVDPEMLTEKWLQSEAPPYPTORVDYGMLYOTRWPLLRFAFAGFRAR 120		
DB	61	SFSAPAGNPYLVDPEMLTEKWLQSEAPPYPTORVDYGMLYOTRWPLLRFAFAGFRAR 120		
QY	121	ASAOQKTELEAFIEAERFWLEDYALFWALKTRFDGKPNWNSPELRDREPAALARAREEL 180		
DB	121	ASAOQKTELEAFIEAERFWLEDYALFWALKTRFDGKPNWNSPELRDREPAALARAREEL 180		
QY	181	AEVALYEWIOWLFYLENGQTKAYAESKGIQIGDMPIFVAFDSSDVWANPOYFYLEADG 240		
DB	181	AEVALYEWIOWLFYLENGQTKAYAESKGIQIGDMPIFVAFDSSDVWANPOYFYLEADG 240		
QY	241	NPTVAGVPRDYFSETGQLWGNPLYRWDMERDPAWCIARIQSLKQCHLVRIHDHFRGF 300		





Db 2 GATSYGSPYQSFSAFAGNTHFIDLDILVEQCLLEASDLEGDFGSDASEVDYAKIYYAR 61  
Qy 107 WPLLRAFAFRASADQKTRLEAFTEAERFWLEDYALFMALKTRFDGKPMNEW-SPDL 165  
Db 62 RPLLEKAVGFEVGVKDD---FEKFAQDNQSWLELFAEYMAIKHEFDNLAWTEWPDADA 118  
Qy 166 RDREPAALAREBELAEVALYEWIOWLFYLEWQOTKAYAESKGIQIIGDMPFVAFDSS 225  
Db 119 RARKASALEVREQLADKLVVHRVYQYFFQWQKLKAYANDHIEIVGDMPIYVAEDSS 178  
Qy 226 DVANPOYFLEADGNTVAGVPRDYFSETGOLWGNPLRYMDVMDRNFAMCIARIQS 285  
Db 179 DMWNPFLFTDNGKATCIAGCPDDEFSTGOLWGNPIYDWEAMDKGKMWIERLES 238  
Qy 286 LKQCHLVRIIDHFRGFAYWEVPGRPNAVGRWKAPEKLFPAVRAQLSDAPITIAEDLG 345  
Db 239 FKIIDIVRIIDHFRGFESYWEIPAGSDTAAPGEWKGFGYKLFPAVKEELGELNIIABDLG 298  
Qy 346 VITPEVALRDGFGPGMKILOAFSGEDNAF-LPHNYPAGHNVVYSGTHDNDTTLGW 404  
Db 299 FMTDEVIELRERTGFGPGMKILOAFNPEDESIDSPLAPA--NSVMYTGTHDNTVLGWY 356  
Qy 405 RTAPEAERAPMAYLARYGIRCLSEYB-VAGALIELAFKSPAKLAIIVPLQDVLGLGPEAR 463  
Db 357 RN--EIDDA--TREYMARYTNR--KEYETVVHMLRTVFSVSMATIMQDLELDEAAR 411  
Qy 464 MNPFGLDGNWARYAEGDLERGLAAGLALAEASQR 500  
Db 412 MNFPSTLGGNWSRMTEDQLTAVPEGLDLTTIYRR 448

## RESULT 4

US-10-424-599-170978  
; Sequence 170978, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 170978  
; LENGTH: 567  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(567)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_125408C.1.pap  
US-10-424-599-170978

Query Match 36.4%; Score 989; DB 4; Length 567;  
Best Local Similarity 44.2%; Pred. No. 1.2e-83;  
Matches 214; Conservative 71; Mismatches 181; Indels 18; Gaps 8;

Qy 2 QLORAFGLLHPTSPFGRMGIGALGREAFRLWDLADAGARMWQVLPGLP-----TSYGDS 57  
Db 67 ELRRRCGLLHPTSPFGRMGIGALGREAFRLWDLADAGARMWQVLPGLP-----TSYGDS 126  
Qy 58 PYCSFSAFAGNPLVDPEMLIEKGLQSEAPPYPTQVDYGLWYQTRWPLRRFAFAG 117  
Db 127 PYSGDQNCNTLLISLEGVDEGLLEKHELPOLDAERNVSLVADKDLITKA-AER 185  
Qy 118 RARASADQKTRLEAFTEAERF--WLEDYALFMALKTRFDGKPMNEWSPDLRDEPAALAR 175  
Db 186 LISSEGLKTOLEFRDPPDISWLEDAAYFAAIDSLNTISWYNWPEPLRNHVLVAED 245

Qy 176 AREELAEVALYEWIOWLFYLEWQOTKAYAESKGIQIIGDMPFVAFDSSDVWANPOVEY 235  
Db 246 IYQKDDFINVFAQOFLFORQKQVHSAQSKGISIMGDMPYIVGVHSDVWANKQKFL 305  
Qy 236 LEADGNTVAGVPRDYFSETGOLWGNPLRYMDVMDRNFAMCIARIQSILKQCHLVRIID 295  
Db 306 LNRKGFPLLVSGVPPDAFSETGOLWGSPLYDWKAMEKGYSWWVRRIRRAQNLDFEPRID 365  
Qy 296 HFRGFAYWEVPGRPNAVGRWKAPEKLFPAVRAQLSDAPITIAEDLGVIITPEVALR 355  
Db 366 HFRGFAYWAVPSEAKVAMLGK--XGFGISLFTTIFRAVGRINIIAEDLGVIITPEVALR 424  
Qy 356 DGFPGFGMKILOAFSGE--DNAFLPHNYPAGHNVVYSGTHDNDTTLGWERTPEAERAF 414  
Db 425 RSTGAPGMAVLQFGFGDANNPHLPHNEC--NQVYTGTHDNDTTLGWERTPEAERAF 482  
Qy 415 MRAYLARYGIRCLSE--YEVAGALIELAFKSPAKLAIIVPLQDVLGLGPEARMPFGRGLDN 473  
Db 483 VLSYLS-----LNEGDEISWALIQRLASVAQTAIIPMDVLGLGNSARMNIPATQFGN 536  
Qy 474 WAWR 477  
Db 537 WGRW 540

## RESULT 5

US-10-425-114-57203  
; Sequence 57203, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 57203  
; LENGTH: 486  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMFLB73152G04\_FLI.pap  
US-10-425-114-57203

Query Match 33.0%; Score 894.5; DB 4; Length 486;  
Best Local Similarity 42.0%; Pred. No. 7.3e-75;  
Matches 196; Conservative 63; Mismatches 193; Indels 15; Gaps 7;

Qy 18 GRWGTGALGREAFRLWDLADAGARMWQVLP---GPTSYGD--SPYQSFSAFAGNPLVD 73  
Db 1 GPHGIGDLGDEALAFHLWLRDAGCTLMQVLPVPPGRTAGDEGSPYSGDQNCNTLLS 60  
Qy 74 PEMLEKGLQSEAPPYPTQVDYGLWYQTRWPLRRFAFAGFRARASADQKTRLEAFI 133  
Db 61 LSELVKGDLMSDELDPDPTVEYFEDTVANLKEPLIAKA-AERLLSHGELATQYDCFK 119  
Qy 134 EAERF--WLEDYALFMALKTRFDGKPMNEWSPDLRDEPAALAREBELAEVALYEWIQ 191  
Db 120 KNPNTSGWLEDAALFAAIDKSINAFSWYEWPEPLKRNHLGALEDIYEKQKDFIETEMAQ 179  
Qy 192 WLFYLEWQOTKAYAESKGIQIIGDMPFVAFDSSDVWANPOVEYFLEADGNTVAGVPRD 251  
Db 180 FUFQWQIRKYAQKLGISIMGDMPYIVGVHSDVWANKSFLLDKNGFPPTVSGVPPD 239  
Qy 252 YFSETGOLWGNPLRYMDVMDRNFAMCIARIQSILKQCHLVRIIDHFRGFAYWEVPGFRP 311

Db 240 AFSETGQNSPLDYKWKMEADCFSWWIKRIKRALDLYDEFRIDHFRGLAGFVAVPSDAK 299

Qy 312 NAVGRVWKAPEKLFVAVRAQSDAPIIAEDLGVIITPEVEALRDGFGFGKILQFAP- 370

Db 300 VALVGSWRAGPRNSFFDTLFAVGRIDIIAEDLGVIITPEVEALRDGFGFGKILQFAPG 359

Qy 371 SGEDNAELPHNYPAHGNVYVSGTHDNDTTLGWERTAPEAERAFMRAYLARYGIRCLSEY 430

Db 360 GGSNDPLPHNHEM--DOVVTGTHDNDTTLGWMONLPEEKKIVIKLPE-----AENI 412

Qy 431 EVAGALTEAFKSPAKLAIPLQDVLGLGPEARNFPGRLGDNMAWR 477

Db 413 DITWILITRALSSVARTSVTMDIILGDSARNTTATQKGNWR 459

RESULT 6

US-10-336-597-5

Sequence 5, Application US/10336597

Publication No. US2003015002A1

GENERAL INFORMATION:

APPLICANT: KREBBERS, ENNO

PEARLSTEIN, RICHARD W.

BROGLIE, KAREN E.

TITLE OF INVENTION: PLANT 4-(GLUCANOTRANSFERASE

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY

STREET: 1007 MARKET STREET

CITY: WILMINGTON

STATE: DELAWARE

COUNTRY: UNITED STATES OF AMERICA

ZIP: 19898

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.50 INCH DISKETTE

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: MICROSOFT WINDOWS 95

SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (VER. 7.0)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/336,597

FILING DATE: 03-Jan-2003

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/403,332

FILING DATE: 19-Oct-1999

ATTORNEY/AGENT INFORMATION:

NAME: MAJARIAN, WILLIAM R.

REGISTRATION NUMBER: 41,173

REFERENCE/DOCKET NUMBER: BB-1101-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 302-992-4926

TELEFAX: 302-773-0164

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 298 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-10-336-597-5

Query Match 28.6%; Score 777.5; DB 4; Length 298;

Best Local Similarity 53.2%; Pred. No. 3.6e-64;

Matches 156; Conservative 39; Mismatches 89; Indels 9; Gaps 6;

Qy 210 IQIGDMPIFVAFDSSDVWNPQFYLEADGNPTVAVGPRDYSETGQMGNPLRYMDV 269

Db 3 IEIVGDMPIYVAEDSSDWANPHLFTKDVNGKATCIAGCPDPEFSVTGQMGNPLIYDWEA 62

Qy 270 MERDNFANCIARIOSLQCHLVRIIDHFRGFAYWEVFGGRNAVGEGRWKAPEKLFPA 329

Db 63 MDKQYKWKIERESFKIYDIVRIDHFRGFESYWEIPAGSDTAPGEWKGPGYKLFPA 122

Qy 330 VRAQSDAPIIAEDLGVIITPEVEALRDGFGFGKILQFAPGSDNAF-LPHNYPAGNV 388

Db 123 VKHEELGELNIIAEDLGFWTDEVIELRERTGFGKILQFAPNPEDESIDSHLAPA--NS 180

Qy 389 VVYSGTHDNDTTLGWERTAPEAERAFMRAYLARYGIRCLSEYE-VAGALIELAFKSPAKL 447

Db 181 VMYTGHNDNTVLGWYRN--EIDDA-TREYMARYNR--KEYETVVHAMURTVFSSVSFM 235

Qy 448 AIVPLQDVLGLGPEARNFPGRLGDNMAWRVYAEGLDEPLGLAAGLRALAEASQR 500

Db 236 AIATMODLELEDEARNFESTLGGNWSNRWTEQDTPAVEEGLLDITTIYR 288

RESULT 7

US-10-425-115-325328

Sequence 325328, Application US/10425115

Publication No. US20040214272A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

FILE OF INVENTION: Plants

FILE REFERENCE: 38-21(53222)B

CURRENT APPLICATION NUMBER: US/10/425,115

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 369326

SEQ ID NO 325328

LENGTH: 486

TYPE: PRT

ORGANISM: Zea mays

FEATURE:

NAME/KEY: unsure

LOCATION: (1)..(486)

OTHER INFORMATION: unsure at all Xaa locations

FEATURE:

OTHER INFORMATION: Clone ID: MRT4577\_59770C.1.pap

US-10-425-115-325328

Query Match 28.3%; Score 767.5; DB 4; Length 486;

Best Local Similarity 37.3%; Pred. No. 6.2e-63;

Matches 185; Conservative 58; Mismatches 180; Indels 73; Gaps 11;

Qy 18 GRWIGNAGREAREFLDMLADAGARWQVLP---GPTSYGD-SPYQSFSAGAGNPYLVD 73

Db 1 GPHGIDGLDEALAPLHLWRDAGCTLMQVLPVPPGRTAGDGSFYSGQDANCNTLLS 60

Qy 74 PEMLEIKGWLQSEAPPPYPTQRYDYGLWYQTRWPLLRAPAGFRARASAOQKTRLEAFI 133

Db 61 LEELVKDGLMSDELDPDVTVEYFPTVANLKEPLIACA----- 100

Qy 134 EAERFWL-----EDYALFMALKTRFDGKPNWEW--SPELRDREPAALARA 176

Db 101 -AERLLSXESSEHNMTASKKIQIYLLVGLTGRQFSLXFRMXDNPHLPH----- 150

Qy 177 REELAEVALYEWIQLWLFYLEWQ-----TKAYA-ESKGIQIIGDMPFVAF 222

Db 151 NHENDQVYTYXEHMITIQLAGGKIYORKKITWAKYPRKYAXKTILISINGDMPLYAGY 210

Qy 223 DSSDVWNPQFYLEADGNPTVAVGPRDYSETGQMGNPLRYMDVNERDNFANCIARI 282

Db 211 HSAYVWNRKSFLLDKNGFTTFVSGVPDAFSETGQMGNPLRYMDWKADEACFSWIKRI 270

Qy 283 RQSLQCHLVRIIDHFRGFAYWEVFGGRNAVGEGRWKAPEKLFVAVRAQSDAPIAE 342

Db 271 KRALDLYDEFRIDHFRGLAGFWAVPSDAKVAVGSRWAGPRNSPFDTLFKAVGRIDIIE 330

Qy 343 DLGVIITPEVEALRDGFGFGKILQFAP-SCEDNAFLPHNYPAGNVVYVSGTHDNDTTL 401

Db 331 DLGVIITPEVEALRDGFGFGKILQFAPGNAVLPQFAPGGSDNPHLPHNHEM--DQVVTGTHDNDTVL 388

Qy 402 GWERTAPEAERAFMRAYLARYGIRCLSEYEYVAGALIELAFKSPAKLAIPLQDVLGLPE 461

Db 389 GWMQNLFEKIKVILPE-----AENIDITWLTALSSVARTSVMTQDILGLDSS 443  
QY 462 ARMMFPGRLGDNWAWR 477  
Db 444 ARMTPTATQKGNWWR 459

RESULT 8  
US-10-336-597-6  
; Sequence 6, Application US/10336597  
; Publication No. US20030150021A1  
; GENERAL INFORMATION:  
; APPLICANT: BROGLIE, KAREN E.  
; PEARLSTEIN, RICHARD W.  
; TITLE OF INVENTION: PLANT 4-(-GLUCANOTRANSFERASE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.50 INCH DISKETTE  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (VER. 7.0)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/336,597  
; FILING DATE: 03-Jan-2003  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/403,332  
; FILING DATE: 19-Oct-1999  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MAJARIAN, WILLIAM R.  
; REGISTRATION NUMBER: 41,173  
; REFERENCE/DOCKET NUMBER: BB-1101-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-992-4926  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 291 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-336-597-6

Query Match 25.6%; Score 694; DB 4; Length 291;  
Best Local Similarity 50.2%; Pred. No. 2.4e-56;  
Matches 136; Conservative 40; Mismatches 87; Indels 8; Gaps 3;

QY 208 KGQITGDMPIFVAFDSSDVWANPQVYFLEADGNPTVAGVPRDYFSETGQLMGNPLYRW 267  
Db 1 KGISMGDMPIYGVHSADWANKQFLNKRKGFPLIVSGVPPDAFSETGQLWGSPLYDW 60

QY 268 DVMERDNFACIARIOSLKQCHLVRIIDHPRGFEAYWEVFPGRNVAEKGKFL 327  
Db 61 KAMEKGFSSWVRRIQRATDLDFEIDHPRGFEAYWEVFPGRNVAEKGKFL 120

QY 328 AAVRAQLSDAPIAEDLVITPEVEALRDGFGFGPKILOFAP-SCEDNAFLPHNPAHC 386  
Db 121 DAILOAVGKINIIAEDLVITPEVEALRDGFGFGPKILOFAP-SCEDNAFLPHN- -EQ 178

QY 387 NVVYVSTHNDTTLGWFTAPAEAFMAYLARYGIRCLSEYVAGALIELAFKSPAK 446  
Db 179 NQVVYTGTHNDTIRGWDTLPQEEKSNVLKLSN-----IEEEISRGILIEGAVSSVAR 233

QY 447 LAIVPLQDVLGLGPEARMNFPGRGLGDNWAWR 477  
Db 234 IAIIPMQDVLGLGSDSRMNIIPATQFGNWSWR 264

RESULT 9  
US-10-437-963-138565  
; Sequence 138565, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 138565  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_39941C.1.pap  
US-10-437-963-138565

Query Match 17.6%; Score 477; DB 4; Length 243;  
Best Local Similarity 45.3%; Pred. No. 4.6e-36;  
Matches 101; Conservative 25; Mismatches 89; Indels 8; Gaps 3;

QY 256 TGOLGNPLYRVDVNERDNFACIARIOSLKQCHLVRIIDHPRGFEAYWEVFPGRNVAE 315  
Db 1 TQOLMNSPLYDWKAMEAGGFEMWIKRINRALLDYDFRIDHFRGLAGFWAVPSESKVALV 60

QY 316 GRWVKAPGEKLFAAVRAQLSDAPIAEDLVITPEVEALRDGFGFGPKILOFAP-SCGD 374  
Db 61 GSWRAGPRNAFFDALFKAVGRINIIAEDLVITPEVEALRDGFGFGPKILOFAP-SCGD 120

QY 375 NAFPLPHNPAHCNVVYVSTHNDTTLGWFTAPAEAFMAYLARYGIRCLSEYVAG 434  
Db 121 NPHLPHN- -EFDQVYVYTGTHNDTIVGWQTLPEEEKQTVFKYLP- -ANRTEISW 173

QY 435 ALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRGLGDNWAWR 477  
Db 174 ALITAAALSSVARTSVMTQDILGLDSSARMTPTATQKGNWWR 216

RESULT 10  
US-09-841-132-518  
; Sequence 518, Application US/09841132  
; Patent No. US20020061848A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Probst, Peter  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; FILE REFERENCE: 210121.469C8  
; CURRENT APPLICATION NUMBER: US/09/841,132  
; CURRENT FILING DATE: 2001-04-23  
; NUMBER OF SEQ ID NOS: 599  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 518  
; LENGTH: 526  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae

US-09-841-132-518  
Query Match 16.6%; Score 451; DB 3; Length 526;  
Best Local Similarity 26.8%; Pred. No. 3.6e-33;  
Matches 131; Conservative 87; Mismatches 212; Indels 58; Gaps 15;  
Qy 21 GIGALGREAERFLD-----WLADAGARWQVPLGPTSYGDSYQSFSAFAGNPYLVDP 74  
Db |||||  
41 GIG-----BFLDLIPLISWCQKQGFVQIQLLPLNDTGEDTSPYNSISSVALNPLFLSL 93  
Qy |||||  
75 EML-----IEKWLQSEAPPYPTQRYDYGWLYQTRWPLLRRAAFAGFRASAQDK 126  
Db |||||  
94 SSLNIDTIPEVAKLQDMHEL---CSTPSYSYQVKEKWAFLREYQKC-CXSSLEGN 149  
Qy |||||  
127 TRLEAFIEAERFWLEDYALFMALKTRFDGKPNWNSPELRDREPAALARAEEELAEVAL 186  
Db |||||  
150 SNFSEFLESERYWLYPYGTFRAIKHHMGEPIINNPKSLTDQE--NFPDLTKKHFDEVLF 207  
Qy |||||  
187 YEWIOWLFLEWGTQKAYAESKGIQIIGDMPIFVAFDSSDVWNPVPOFYFYLEADGNPTVA 246  
Db |||||  
208 FSYLQFLCYOQLCEVAYADQHHVLLKGDLPILISKDCDVWYFRDYF-----SSRSV 261  
Qy |||||  
247 GVPDYFSETQOLGNPLRYDVMERNFANCIARIQSLKQCHLVRIIDHFRGFEAYWEV 306  
Db |||||  
262 GAPPDLYNSEGQWHLPIYNSQLAKDDYIWWKRLRYAQNFYSVYRLDHIIGFFRLWI- 320  
Qy |||||  
307 PFGRPNAVEGRWK-----APGEKLFPAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359  
Db |||||  
321 ---WDSSGRGRFIPDNPKDYIKQGTIELSTMLGASSMLP-IGEDLGIIPQDVKTTLHLG 376  
Qy |||||  
360 FPGMKILOFASGE-DNAFLPHNYPAGNVVVVSGTHDNDTTLGWFTAPAEAFNRAY 418  
Db |||||  
377 ICGTRIPRWRNWSDSAFIPLKDYNPVLTLS-THDSDTFAQWMLNSPKAKQFAKFL 435  
Qy |||||  
419 LARYGIRCLSEVAVAGALIELAFKSPAKLAIVPLQDVGLGPE-----ARNFPGRL 470  
Db |||||  
436 HLPFQKLTETQI--DILKLSHESAFIHNFNDYLALCPDLVSKNLOQERINTPTGI 493  
Qy |||||  
471 G-DNNAWR 477  
Db |||||  
494 SKKNWSYR 501  
RESULT 11  
US-10-289-762-347  
Sequence 347, Application US/10289762  
Publication No. US20040006218A1  
GENERAL INFORMATION:  
APPLICANT: Griffais, R.  
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection  
FILE REFERENCE: 9710-003-999  
CURRENT APPLICATION NUMBER: US/10/289,762  
CURRENT FILING DATE: 2003-03-27  
NUMBER OF SEQ ID NOS: 6849  
SEQ ID NO 347  
LENGTH: 526  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
US-10-289-762-347  
Query Match 16.6%; Score 451; DB 4; Length 526;  
Best Local Similarity 26.8%; Pred. No. 3.6e-33;  
Matches 131; Conservative 87; Mismatches 212; Indels 58; Gaps 15;  
Qy 21 GIGALGREAERFLD-----WLADAGARWQVPLGPTSYGDSYQSFSAFAGNPYLVDP 74  
Db |||||  
41 GIG-----BFLDLIPLISWCQKQGFVQIQLLPLNDTGEDTSPYNSISSVALNPLFLSL 93  
Qy |||||  
75 EML-----IEKWLQSEAPPYPTQRYDYGWLYQTRWPLLRRAAFAGFRASAQDK 126  
Db |||||  
94 SSLNIDTIPEVAKLQDMHEL---CSTPSYSYQVKEKWAFLREYQKC-CXSSLEGN 149

Qy 127 TRLEAFIEAERFWLEDYALFMALKTRFDGKPNWNSPELRDREPAALARAEEELAEVAL 186  
Db |||||  
150 SNFSEFLESERYWLYPYGTFRAIKHHMGEPIINNPKSLTDQE--NFPDLTKKHFDEVLF 207  
Qy |||||  
187 YEWIOWLFLEWGTQKAYAESKGIQIIGDMPIFVAFDSSDVWNPVPOFYFYLEADGNPTVA 246  
Db |||||  
208 FSYLQFLCYOQLCEVAYADQHHVLLKGDLPILISKDCDVWYFRDYF-----SSRSV 261  
Qy |||||  
247 GVPDYFSETQOLGNPLRYDVMERNFANCIARIQSLKQCHLVRIIDHFRGFEAYWEV 306  
Db |||||  
262 GAPPDLYNSEGQWHLPIYNSQLAKDDYIWWKRLRYAQNFYSVYRLDHIIGFFRLWI- 320  
Qy |||||  
307 PFGRPNAVEGRWK-----APGEKLFPAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359  
Db |||||  
321 ---WDSSGRGRFIPDNPKDYIKQGTIELSTMLGASSMLP-IGEDLGIIPQDVKTTLHLG 376  
Qy |||||  
360 FPGMKILOFASGE-DNAFLPHNYPAGNVVVVSGTHDNDTTLGWFTAPAEAFNRAY 418  
Db |||||  
377 ICGTRIPRWRNWSDSAFIPLKDYNPVLTLS-THDSDTFAQWMLNSPKAKQFAKFL 435  
Qy |||||  
419 LARYGIRCLSEVAVAGALIELAFKSPAKLAIVPLQDVGLGPE-----ARNFPGRL 470  
Db |||||  
436 HLPFQKLTETQI--DILKLSHESAFIHNFNDYLALCPDLVSKNLOQERINTPTGI 493  
Qy |||||  
471 G-DNNAWR 477  
Db |||||  
494 SKKNWSYR 501  
RESULT 12  
US-10-872-155-518  
Sequence 518, Application US/10872155  
Publication No. US20040234536A1  
GENERAL INFORMATION:  
APPLICANT: Bhatia, Ajay  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Probst, Peter  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND DIAGNOSIS OF CHLAMYDIAL INFECTION  
FILE REFERENCE: 210121.469C10  
CURRENT APPLICATION NUMBER: US/10/872,155  
CURRENT FILING DATE: 2004-06-18  
PRIOR APPLICATION NUMBER: 09/620,412  
PRIOR FILING DATE: 2000-07-20  
PRIOR APPLICATION NUMBER: 09/598,419  
PRIOR FILING DATE: 2000-06-20  
PRIOR APPLICATION NUMBER: 09/556,877  
PRIOR FILING DATE: 2000-04-19  
PRIOR APPLICATION NUMBER: 09/454,684  
PRIOR FILING DATE: 1999-12-03  
PRIOR APPLICATION NUMBER: 09/426,571  
PRIOR FILING DATE: 1999-10-22  
PRIOR APPLICATION NUMBER: 09/410,568  
PRIOR FILING DATE: 1999-10-01  
PRIOR APPLICATION NUMBER: 09/288,594  
PRIOR FILING DATE: 1999-04-08  
PRIOR APPLICATION NUMBER: 09/208,277  
PRIOR FILING DATE: 1998-12-08  
NUMBER OF SEQ ID NOS: 599  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 518  
LENGTH: 526  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
US-10-872-155-518  
Query Match 16.6%; Score 451; DB 5; Length 526;  
Best Local Similarity 26.8%; Pred. No. 3.6e-33;  
Matches 131; Conservative 87; Mismatches 212; Indels 58; Gaps 15;  
Qy 21 GIGALGREAERFLD-----WLADAGARWQVPLGPTSYGDSYQSFSAFAGNPYLVDP 74  
Db |||||

Db 41 GIG-----EFLDLIPLISWCKQGFVSIQLPLNDTGBDTSYNGSSVALNPLFLSL 93  
QY 75 EML-----IEKGWLEQSEAPPYPTQRTVDYGLWYQTRWPLLRARAFAGFRASADK 126  
Db 94 SSLPNIDTIEVAKLQDMHEL-----CSTPSVSYTVKKEKWAFLREYYQKC-CKSSLEGN 149  
QY 127 TRLEAFTEARFWLEDYALFWALKTRFDGKPNWNSPELDRDREPAALARAREELAEVAL 186  
Db 150 SNFSEFLESRYWLYPGTFRAIKHMHGEPINNWPKSLTDQE--NFPDLTKKHFDEVLF 207  
QY 187 YEWIOWLFYLEWGQTKAYASKGQIIGDMPIFVAFDSSDVWNPQVYFLEADGNPTVA 246  
Db 208 FSVLQFLCYQCLCEKAYADQHHVLLKGLDPLITSKSDCVWYFRDYF-----SSRSV 261  
QY 247 GVPDYFSEGTQWGNPLRWDMERONFANCIAIROSLOKCHLVRIIDHFRGEAYEV 306  
Db 262 GAPDPLNSGQWHLPIYFNSQAKDDYIWKERLYAQNFSVYKLDHIIGFFRLWI- 320  
QY 307 PFGPNAVEGRWVK-----APEKLFVAAVRAQLSDAPIIAEDLGVTIPEVALRDGFG 359  
Db 321 ---WSSGRCRFPDPKDIKQTEILSTMLGASSMLP-IGEDLGIIPQDVKTTLHLG 376  
QY 360 FPGMKILOPAPSGE-DNAPLPHNYPAHNGVYVSGTHDNDTTLGWFTAPAEARAFRAY 418  
Db 377 ICGTRIPRWERNWESDAFIPLKDYNPLSVTTLT--THDSDTFAQWMLNSPKEAKQFAKFL 435  
QY 419 LARYGIRCLSEYEVAGALIELAFKSPAKLAIVLODVILGLPE-----ARNNPPGRL 470  
Db 436 HLPFQKTLTTETQI--DILKLSHESAIPIHNFNDYALCPDLVSKNQLQRINTPGTI 493  
QY 471 G-DNMAWR 477  
Db 494 SKXWSYR 501

RESULT 13  
US-09-841-132-588  
; Sequence 588, Application US/09841132  
; Patent No. US20020061848A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Probat, Peter  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C8  
; CURRENT APPLICATION NUMBER: US/09/841,132  
; NUMBER OF SEQ ID NOS: 599  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 588  
; LENGTH: 527  
; TYPE: PRT  
; ORGANISM: C. Trachomatis D serovar  
US-09-841-132-588

Query Match 15.9%; Score 430.5; DB 3; Length 527;  
Best Local Similarity 25.9%; Pred. No. 3.1e-31;  
Matches 126; Conservative 89; Mismatches 233; Indels 39; Gaps 12;  
QY 33 LNLADAGARWQVLPGLGTSYGDSPYQSFSAFAGNPNYLDPEMLIEKWLGESEAPPY 92  
Db 56 IDWCISGQILQILPINDTSGSSPYNSISSIALNPLHLSIALPYKEEVPAAETRIRE 115  
QY 93 PTQRTVDYGLWYQTRWPLLRARAFAGFRASAQ-----DKTRLEAFIEARFWLEDYALFWA 148  
Db 116 MOOLSOLPOVHYEKVRKMRDFFQYRYVCKQKLTDPDFYAFCEQEKYHLHPALFRS 175  
QY 149 LKTRFDGKPNWNSPELDRDREPAALARAREELAEVALYEWIOWLFYLEWGQTKAYAESK 208  
Db 176 IREHLDNLINHWPTTYTDL--SQITEHRTFAEDIQFHSYLOYLFCFQMTQVREHACK 233  
QY 209 GIQIIGDMPIFVAFDSSDVWNPQVYFLEADGNPTVAVGPVDFSETGQLMGNPLRW 268

Db 234 SCLIKGDIPLISKSDCVWFYRHP-----SSSESVGAPPDLYNAEQWHLPICNWK 287  
QY 269 VMERDNFAMCIARIOSLOKCHLVRIIDHFRGFPAYW---EVPGRPNVAGRWKAPCEK 325  
Db 288 TLQODNYLWKEKRLRYAENFYSLYRLDHWVGLFRFWWDESGGCRPEPHDPKNYLAQOD 347  
QY 326 LPAVRAQLSDAPIIAEDLGVTIPEVALRDGFGPKMKILOAFSGEDN-AFLPHNYP 384  
Db 348 ILSHLTSSMLP-IGEDLOTSPDVKRMLESFAVCGTRIPOWERWEGNGATP--FDQ 404  
QY 385 HGNVVVYS-GTHDNDTTLGWFTAPAEARAFMR-----AYLARYGIRCLSEYEVAGALIEL 439  
Db 405 YDPLSVTSLTHDSSLASWKPESQESKLFQAFLGLPYSTSLSHNHT-ILKL 458  
QY 440 AFKSPAKLAIVLODVILGLPE-----ARNNPPGRLG-DNMAWRYAGDLEPGLAAG 490  
Db 459 SHKTSIFRINLINDYALFPOLISKTPIRYERINLEGTISKNNWYVRKPSIEDLSHSK 518  
QY 491 LRALAEA 497  
Db 519 LNSLLER 525

RESULT 14  
US-10-872-155-588  
; Sequence 588, Application US/10872155  
; Publication No. US20040234536A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Probat, Peter  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C10  
; CURRENT APPLICATION NUMBER: US/10/872,155  
; CURRENT FILING DATE: 2004-06-18  
; PRIOR APPLICATION NUMBER: 09/620,412  
; PRIOR FILING DATE: 2000-07-20  
; PRIOR APPLICATION NUMBER: 09/598,419  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/556,877  
; PRIOR FILING DATE: 2000-04-19  
; PRIOR APPLICATION NUMBER: 09/454,684  
; PRIOR FILING DATE: 1999-12-03  
; PRIOR APPLICATION NUMBER: 09/426,571  
; PRIOR FILING DATE: 1999-10-22  
; PRIOR APPLICATION NUMBER: 09/410,568  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: 09/288,594  
; PRIOR FILING DATE: 1999-04-08  
; PRIOR APPLICATION NUMBER: 09/208,277  
; PRIOR FILING DATE: 1998-12-08  
; NUMBER OF SEQ ID NOS: 599  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 588  
; LENGTH: 527  
; TYPE: PRT  
; ORGANISM: C. Trachomatis D serovar  
US-10-872-155-588

Query Match 15.9%; Score 430.5; DB 5; Length 527;  
Best Local Similarity 25.9%; Pred. No. 3.1e-31;  
Matches 126; Conservative 89; Mismatches 233; Indels 39; Gaps 12;  
QY 33 LNLADAGARWQVLPGLGTSYGDSPYQSFSAFAGNPNYLDPEMLIEKWLGESEAPPY 92  
Db 56 IDWCISGQILQILPINDTSGSSPYNSISSIALNPLHLSIALPYKEEVPAAETRIRE 115  
QY 93 PTQRTVDYGLWYQTRWPLLRARAFAGFRASAQ-----DKTRLEAFIEARFWLEDYALFWA 148  
Db 116 MOOLSOLPOVHYEKVRKMRDFFQYRYVCKQKLTDPDFYAFCEQEKYHLHPALFRS 175



```

RESULT 15
US-10-767-701-51019
; Sequence 51019, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 51019
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3480-052-P1-K1-G9.pgp
; US-10-767-701-51019

```

	Query Match	15.2%	Score 412.5	DB 4	Length 179
	Best Local Similarity	49.4%	Pred. No. 3.5e-30		
	Matches	83	Conservative 18	Mismatches 64	Indels 3
					Gaps 2
Qy	253	PSETGOLGNPLDYDWMERNDFANWCARIQSLSKQCHLVDFIDHFRGFEAYNEVPFGRPN	312		
Db	2	PSETGOLNPLDYDWMKMEADGFSWTKRIKRALDLVDFRIDHFRGLAGFWAVPSAKV	61		
Qy	313	AVEGRWKAQCEKLPAAVRAQSDAPIIAEDLGWITTPVEALRDGFGFPFGMKIIQAF-S	371		
Db	62	ALVGSWRAGPRNSFPDALFKAVGRINIAEDLGWITDVVQLRKSIGAPGNVAVQAFGG	121		
Qy	372	GEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFTAPAEAEAFWRAYL	419		
Db	122	GSDNPHLPNHHEM--DOVVYTGTHDNDTVLGGWQNLPEEEKIKVIKYL	167		

Search completed: January 10, 2006, 14:24:33  
Job time : 116 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2006, 14:13:34 ; Search time 29 Seconds  
(without alignments)  
146.735 Million cell updates/sec

Title: US-10-618-976-2  
Perfect score: 2714  
Sequence: 1 MQLQRAFGILLHTSPGWR.....DLEPGLAAGLALASQRA 501

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 61141 seqs, 8493638 residues

Total number of hits satisfying chosen parameters: 61141

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New:  
1: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pap:  
2: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB.pap:  
3: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pap:  
4: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pap:  
5: /cgn2\_6/prodata/1/pubpaa/US09\_NEW\_PUB.pap:  
6: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB.pap:  
7: /cgn2\_6/prodata/1/pubpaa/US11\_NEW\_PUB.pap:  
8: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	3.7	658	US-10-873-528-17	Sequence 17, Appl
2	100	3.7	677	US-10-873-528-155	Sequence 155, Appl
3	90	3.3	2644	US-10-770-728-45	Sequence 45, Appl
4	89	3.3	633	US-10-467-657-730	Sequence 730, Appl
5	89	3.3	1766	US-11-075-185-10	Sequence 10, Appl
6	88	3.2	620	US-11-055-822-460	Sequence 460, Appl
7	88	3.2	620	US-11-055-822-702	Sequence 702, Appl
8	87	3.2	538	US-11-167-856-20	Sequence 20, Appl
9	86	3.2	512	US-11-012-762-64	Sequence 64, Appl
10	85.5	3.2	695	US-11-038-284-34	Sequence 34, Appl
11	85	3.1	6893	US-11-205-109-14	Sequence 14, Appl
12	84.5	3.1	355	US-10-454-437-102	Sequence 102, Appl
13	84.5	3.1	555	US-10-454-437-100	Sequence 100, Appl
14	84	3.1	379	US-10-858-730-22	Sequence 22, Appl
15	82	3.0	715	US-10-467-657-70	Sequence 70, Appl
16	82	3.0	715	US-10-467-657-3672	Sequence 3672, Ap
17	81	3.0	294	US-10-467-657-934	Sequence 934, Appl
18	80	2.9	395	US-11-075-185-13	Sequence 13, Appl
19	80	2.9	402	US-10-454-437-282	Sequence 282, Appl
20	80	2.9	486	US-10-131-828A-58	Sequence 58, Appl
21	80	2.9	615	US-10-821-234-873	Sequence 873, Appl
22	80	2.9	801	US-10-454-437-278	Sequence 278, Appl
23	79.5	2.9	554	US-10-850-816-6	Sequence 6, Appl
24	79.5	2.9	582	US-11-090-439-58	Sequence 58, Appl
25	79.5	2.9	625	US-11-055-822-522	Sequence 522, Appl

26	79	2.9	561	6	US-10-454-437-330	Sequence 330, App
27	79	2.9	7968	7	US-11-143-980-49	Sequence 49, Appl
28	78.5	2.9	379	6	US-10-858-730-285	Sequence 285, Appl
29	78.5	2.9	589	6	US-10-821-234-1687	Sequence 1687, Ap
30	78	2.9	496	7	US-11-186-284-103	Sequence 103, Appl
31	77.5	2.9	379	6	US-10-858-730-24	Sequence 24, Appl
32	77.5	2.9	379	6	US-10-858-730-281	Sequence 281, Appl
33	77.5	2.9	379	6	US-10-858-730-290	Sequence 290, Appl
34	77.5	2.9	600	7	US-11-205-109-39	Sequence 39, Appl
35	77.5	2.9	777	6	US-10-821-234-1658	Sequence 1658, Ap
36	77.5	2.9	4655	6	US-10-995-561-556	Sequence 556, Appl
37	76	2.8	453	6	US-10-878-556A-141	Sequence 141, Appl
38	76	2.8	568	7	US-11-055-822-240	Sequence 240, Appl
39	76	2.8	568	7	US-11-055-822-1134	Sequence 1134, Ap
40	75.5	2.8	239	7	US-11-165-211-38	Sequence 38, Appl
41	75.5	2.8	239	7	US-11-165-226-48	Sequence 48, Appl
42	75.5	2.8	575	6	US-10-467-657-1456	Sequence 1456, Ap
43	75.5	2.8	626	6	US-10-512-184-49	Sequence 49, Appl
44	75.5	2.8	7102	7	US-11-143-980-48	Sequence 48, Appl
45	75	2.8	1892	7	US-11-075-185-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-10-873-528-17  
; Sequence 17, Application US/10873528  
; Publication No. US20050276814A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbial Technics Limited  
; APPLICANT: Gilbert, Christophe FG  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: PWC/P21129W0  
; CURRENT APPLICATION NUMBER: US/10/873,528  
; CURRENT FILING DATE: 2004-06-23  
; PRIOR APPLICATION NUMBER: US/09/769,787  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: GB 9816337.1  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: US 60/125164  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 388  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 658  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-873-528-17

Query Match	3.7%	Score 100;	DB 6;	Length 658;
Best Local Similarity	18.9%;	Pred. No. 0.19;		
Matches 103;	Conservative 63;	Mismatches 155;	Indels 224;	Gaps 28;
QY	22	IGALGREAEERELDLAD-----AGARWQVPLGPTSGDSPVQSFAFA 66		
DB	116	VGATG--AKVIEDWVYDQYDAWFIKADGQAEKEWLOI-----KGRQYI-----FK 161		
QY	67	GNPVLVDPEMLTEKGLQSEAPPYPTQRYVDYGLYQTRWPLLRRAFRAGFRARASQDK 126		
DB	162	SGGYLLTSQW-INQAYVNASG-----KVQOQWLFQKQY-----QSWFYIKENGYADK 209		
QY	137	TRLEAFIE-----AERFWLEDYALFMALKTRFDGK-----PW---158		
DB	210	---EWIPENGHYVYLLKSGYMAANEIWDKESWFLK---FDGKMAEKWVYDSSHSAQDK 264		
QY	159	-----NEWSFELDRPREAALAREELAEVALYEWIOWLFYLEGQTKAYAESKG 209		
DB	265	FKSGGYMTANEW---IWDKESWFLKSGDKTAEK-----EWV-----YDSHSAQWYFKSGG 313		
QY	210	IQIIGDMPFIVAFDSSDYWANPQVYLEADG-----NP 242		

Db 314 YMTANEW-----IMDKESWFLKSDGKIAEKWVYDSSHSAQWYFKSGGYMAKNE 363  
QY 243 TVVA-----GVPRDYFSETGQLWGNPLYRW 267  
Db 364 TVDYGQLGSDGKWLGGKTTNENAAAYQVVPVTANVYDSGKLSYISQGSVVM-----416  
QY 268 DYMERD-----NFWACIARIOSLKOCHLVRIDHFRGFEAYWEVFPGRPNAVEGRWVKA 321  
Db 417 --LDKDKSDKRLAITISGLSYMKTEDQALDASKDFIPIYE-----S 459  
QY 322 PGKLFPAVRAQLSDAPIIA--EDLGVITPEVEALRDGFGPGMKILOQAFSGEDNAFL- 378  
Db 460 DGRHFYHYV-AQNASIPVASHLSDMVEGKKYISA--DGLHFDGFKL-----ENPFLF 508  
QY 379 -----PHNYPAGNVVYSGTHNDNTTLGWFTAPAEAFRAVYLAIRYGIKLSYEVA 433  
Db 509 KDLTEATNYSABELDKVSLNINNSLL-----ENKGATFKEABEHYHINAL--YLLA 559  
QY 434 GALIE 438  
Db 560 HSALE 564

RESULT 2  
US-10-873-528-155  
; Sequence 155, Application US/10873528  
; Publication No. US20050276814A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbial Technics Limited  
; APPLICANT: Gilbert, Christophe FG  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: PWC/P21129WO  
; CURRENT FILING DATE: 2004-06-23  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 388  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 155  
; LENGTH: 677  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-873-528-155

Query Match 3.7%; Score 100; DB 6; Length 677;  
Best Local Similarity 18.9%; Pred. No. 0.2;  
Matches 103; Conservative 63; Mismatches 155; Indels 224; Gaps 28;  
QY 22 IGALGREAERFLDWLAD-----AGARWQVPLPGPTSYGDSFYQSFSAPA 66  
Db 135 VGATG--AKVIEDWVDSQDANFYKADQOHAKEWLOI-----KGDY-----FK 180  
QY 67 GNPYLVDPEMLIEKGMLEQSEAPPPYPTQRYDYGLWLYOTRWPLLRRAFAFRARASAQDK 126  
Db 181 SGGYLLTSQW-INQAYNASSA-----KQVQWGLFDKQY---QSWFYKENGNYADK 228  
QY 127 TRLEAFIE-----AERFWLEDVALFWALKTRFDGK-----PW-- 158  
Db 229 ---EWIPENGHYIYKSGGYMAANEIWDKESFYLK--FDGKMAEKWYDSSHSAQWY 283  
QY 159 -----NEWPELDRPEALARAAREALAEVALYEMIQMLFYLEWQTKAYAESKG 209  
Db 284 FKSGGYNTANEW---IMDKESWFLKSDGKIAEK-----EW-----YDSSHSAQWYFKSGG 332  
QY 210 IQIIGMPIEFAVDSSDWNANPOYFYLEADG-----NP 242  
Db 333 YMTANEW-----IMDKESWFLKSDGKIAEKWVYDSSHSAQWYFKSGGYMAKNE 382

QY 243 TVVA-----GVPRDYFSETGQLWGNPLYRW 267  
Db 383 TVDYGQLGSDGKWLGGKTTNENAAAYQVVPVTANVYDSGKLSYISQGSVVM-----435  
QY 268 DYMERD-----NFWACIARIOSLKOCHLVRIDHFRGFEAYWEVFPGRPNAVEGRWVKA 321  
Db 436 --LDKDKSDKRLAITISGLSYMKTEDQALDASKDFIPIYE-----S 478  
QY 322 PGKLFPAVRAQLSDAPIIA--EDLGVITPEVEALRDGFGPGMKILOQAFSGEDNAFL- 378  
Db 479 DGRHFYHYV-AQNASIPVASHLSDMVEGKKYISA--DGLHFDGFKL-----ENPFLF 527  
QY 379 -----PHNYPAGNVVYSGTHNDNTTLGWFTAPAEAFRAVYLAIRYGIKLSYEVA 433  
Db 528 KDLTEATNYSABELDKVSLNINNSLL-----ENKGATFKEABEHYHINAL--YLLA 578  
QY 434 GALIE 438  
Db 579 HSALE 583

RESULT 3  
US-10-770-726-45  
; Sequence 45, Application US/10770726  
; Publication No. US20050266409A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Brown, Eugene  
; APPLICANT: Liu, Wei  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING, PREVENTING, AND TREATIN  
; FILE REFERENCE: AML01079 (031896-010000)  
; CURRENT APPLICATION NUMBER: US/10/770,726  
; CURRENT FILING DATE: 2004-02-04  
; NUMBER OF SEQ ID NOS: 48640  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 45  
; LENGTH: 2644  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-770-726-45

Query Match 3.3%; Score 90; DB 6; Length 2644;  
Best Local Similarity 20.7%; Pred. No. 10;  
Matches 73; Conservative 39; Mismatches 139; Indels 102; Gaps 16;  
QY 18 GRWIGALGREAERFLDWLADAGARWQVPLPGPTSYGDSFYQSFSAPAGNPYLVDPEML 77  
Db 1805 GQLLSAKKRDITAFYDSLKLVRAB--QIVPLSAAASFERGSYQ-----RGYEVIVRLHML 1857  
QY 78 IEKGMLEQSEAP-----PPYPTQRYDYGLW-----YOTRWPL--LRRAF-----114  
Db 1858 CE---LEHSIKPLFOHSPGDSQEDSLNWWVARLEMTQNSYRAKEPILALRRALLSLNKR 1914  
QY 115 -----AGFRARASAOQDKTRLEAFIEAFERFWLEDVALFWALK 150  
Db 1915 DYNEMVGECWLOSARVARKAGHHQTAYNALLNAGESRLAELVYERAKWLWSKGDVHOALI 1974  
QY 151 TRFDGK-----PNNEWSPE-----LDRPEALARAAREELAE-----EVALYE 188  
Db 1975 VLQKGVELCFPENETPPEGKNMLIHGRAMLVGRFMEETANPESNAIMKKYKDVTAACLP 2034  
QY 189 WQWLFYLE--WGOTKAYAESKGIQIGDMPIFVAFDSSDWMANPOYFYLEADGNPTVA 246  
Db 2035 WEDGHFYLAHYDKLMPWVTDNKKMEKQGLIRIVLHFG---RSLQY-----GNQFIYQ 2085  
QY 247 GVPRDYFSETGOLW---GNPLYRWVYMERDNFANCIARIRQSLKQCHLVRI 296  
Db 2086 SMPRLM-----TLWLDYGYKAYEWKAGRS---RVQWENDLGKINKVITEH 2129

RESULT 4  
US-10-467-657-730



Publication No. US20050260707A1  
GENERAL INFORMATION:  
APPLICANT: Pompejus, Markus  
APPLICANT: Kroger, Burkhard  
APPLICANT: Schröder, Hartwig  
APPLICANT: Zelder, Oskar  
APPLICANT: Haberhauer, Gregor  
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS  
FILE REFERENCE: BGI-121CPCN  
CURRENT APPLICATION NUMBER: US/11/055,822  
CURRENT FILING DATE: 2005-02-11  
PRIOR APPLICATION NUMBER: 09/606,740  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 60/141,031  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: 60/142,101  
PRIOR FILING DATE: 1999-07-02  
PRIOR APPLICATION NUMBER: 60/148,613  
PRIOR FILING DATE: 1999-08-12  
PRIOR APPLICATION NUMBER: 60/187,970  
PRIOR FILING DATE: 2000-03-09  
PRIOR APPLICATION NUMBER: DE 19930476.9  
PRIOR FILING DATE: 1999-07-01  
PRIOR APPLICATION NUMBER: DE 19931415.2  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931418.7  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931419.5  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931420.9  
PRIOR FILING DATE: 1999-07-08  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1158  
SEQ ID NO 702  
LENGTH: 620  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-11-055-822-702

Query Match 3.2%; Score 88; DB 7; Length 620;  
Best Local Similarity 21.4%; Pred. No. 2.2;  
Matches 113; Conservative 73; Mismatches 156; Indels 186; Gaps 35;

QY 7 FGIILLPTSPGPGWIGALGREARFLDLADAGARWQV----LPLGPTSYGDSPPQSF 62  
DB 160 WGVQFHPESIGGFGHQII----KNFLNL----ARTYRWQTEKTIPLSVDS--AAVPETF 210

QY 63 SFA-----GNPYLVDPPEMLEK-----GWL-EQSEAPPPYPTQRVY 99  
DB 211 FAHSHAFWLDDAQSYLGDSGLARTKTHNVGEGDFTWLKEDLAANSVAPGQGR 270

QY 100 GWLYQTRWPLLRAPAFAGRA-----RASA--ODKTRLEAFIEARFWLED 142  
DB 271 GWGYGVGYEL--KAEAGARAHAHTSSLPDAHLIFADRAIAVESDQVRLALGQDE-WFEE 327

QY 143 -----YALFMALKTRFGKPNWNSPELRDREPAALAREELAE-----EV 184  
DB 328 TIKKHLNLVAPRIPASGHLLQVR-DSK--DEYLDKIR-----RAQLITRGESYEI 376

QY 185 ALYEWIQWLFYLEWGOTKAYAESKGIQIIGDMPIFVAFDSSDVWANPOYF--YLEADGNP 242  
DB 377 CLITTKLQ-----GTTDVAPLAAYLALRG-----ANPTAYGAVLQ-LGDT 414

QY 243 TVVAGVPRDYFS--ETGQMGNPY----RWDVNERDNFACIARIR-----QSLKQC 289  
DB 415 SILSSSPERFITIDSAGYVESKPIKTRPRGRTAQEDQ--EIIAELRSNPKRAENLMIV 472

QY 290 HLVRIDHFRGFAYWEVFPGRPNAVEGRWVKAPGEKLF-----AAVRAQLSDAPIAEDL 344  
DB 473 DLVRNDLARG-----ALPTTVK-----TSKLFVDVETATVHQVLS--TVSAEL 513

QY 345 GVITPEVEALRDGF-----GFGMKILQFAFGSDNAPFLPHNYPAGHNVVYSGTHDND 398  
DB 514 GPRSP-IECVRAAPFGSGMTGAPKLRMTWII--DELEAPRG-----IYSG-----556

QY 399 TTGHWFTAPAEARAFNAYLARYGIRCL-----SEYEVAGALIELA 440  
DB 557 -GLGYFSLDGAVIDLSMW-----IRTLVQNNHNVYGVGGALLALS 595

RESULT 7  
US-11-055-822-702  
; Sequence 702, Application US/11055822

Publication No. US20050260707A1  
GENERAL INFORMATION:  
APPLICANT: Pompejus, Markus  
APPLICANT: Kroger, Burkhard  
APPLICANT: Schröder, Hartwig  
APPLICANT: Zelder, Oskar  
APPLICANT: Haberhauer, Gregor  
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS  
FILE REFERENCE: BGI-121CPCN  
CURRENT APPLICATION NUMBER: US/11/055,822  
CURRENT FILING DATE: 2005-02-11  
PRIOR APPLICATION NUMBER: 09/606,740  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 60/141,031  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: 60/142,101  
PRIOR FILING DATE: 1999-07-02  
PRIOR APPLICATION NUMBER: 60/148,613  
PRIOR FILING DATE: 1999-08-12  
PRIOR APPLICATION NUMBER: 60/187,970  
PRIOR FILING DATE: 2000-03-09  
PRIOR APPLICATION NUMBER: DE 19930476.9  
PRIOR FILING DATE: 1999-07-01  
PRIOR APPLICATION NUMBER: DE 19931415.2  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931418.7  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931419.5  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931420.9  
PRIOR FILING DATE: 1999-07-08  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1158  
SEQ ID NO 702  
LENGTH: 620  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-11-055-822-702

Query Match 3.2%; Score 88; DB 7; Length 620;  
Best Local Similarity 21.4%; Pred. No. 2.2;  
Matches 113; Conservative 73; Mismatches 156; Indels 186; Gaps 35;

QY 7 FGIILLPTSPGPGWIGALGREARFLDLADAGARWQV----LPLGPTSYGDSPPQSF 62  
DB 160 WGVQFHPESIGGFGHQII----KNFLNL----ARTYRWQTEKTIPLSVDS--AAVPETF 210

QY 63 SFA-----GNPYLVDPPEMLEK-----GWL-EQSEAPPPYPTQRVY 99  
DB 211 FAHSHAFWLDDAQSYLGDSGLARTKTHNVGEGDFTWLKEDLAANSVAPGQGR 270

QY 100 GWLYQTRWPLLRAPAFAGRA-----RASA--ODKTRLEAFIEARFWLED 142  
DB 271 GWGYGVGYEL--KAEAGARAHAHTSSLPDAHLIFADRAIAVESDQVRLALGQDE-WFEE 327

QY 143 -----YALFMALKTRFGKPNWNSPELRDREPAALAREELAE-----EV 184  
DB 328 TIKKHLNLVAPRIPASGHLLQVR-DSK--DEYLDKIR-----RAQLITRGESYEI 376

QY 185 ALYEWIQWLFYLEWGOTKAYAESKGIQIIGDMPIFVAFDSSDVWANPOYF--YLEADGNP 242  
DB 377 CLITTKLQ-----GTTDVAPLAAYLALRG-----ANPTAYGAVLQ-LGDT 414

QY 243 TVVAGVPRDYFS--ETGQMGNPY----RWDVNERDNFACIARIR-----QSLKQC 289  
DB 415 SILSSSPERFITIDSAGYVESKPIKTRPRGRTAQEDQ--EIIAELRSNPKRAENLMIV 472

QY 290 HLVRIDHFRGFAYWEVFPGRPNAVEGRWVKAPGEKLF-----AAVRAQLSDAPIAEDL 344  
DB 473 DLVRNDLARG-----ALPTTVK-----TSKLFVDVETATVHQVLS--TVSAEL 513

QY 345 GVITPEVEALRDGF-----GFGMKILQFAFGSDNAPFLPHNYPAGHNVVYSGTHDND 398  
DB 514 GPRSP-IECVRAAPFGSGMTGAPKLRMTWII--DELEAPRG-----IYSG-----556

QY 399 TTGHWFTAPAEARAFNAYLARYGIRCL-----SEYEVAGALIELA 440  
DB 557 -GLGYFSLDGAVIDLSMW-----IRTLVQNNHNVYGVGGALLALS 595

RESULT 7  
US-11-055-822-702  
; Sequence 702, Application US/11055822

Db 514 GPRSP-IECVRAFPFGSMTGAPKRTWMEII--DELEAARG-----IYSG-----556  
QY 399 TTIGWFTAREAEARAFRAYLARIGIRCL-----SEYEVAGALIELA 440  
Db 557 -GLGYFSLDGAVIDLSMV-----IRTLVIQNNHVEYGVGGALLALS 595

## RESULT 8

US-11-167-856-20  
; Sequence 20, Application US/11167856  
; Publication No. US20050268352A1  
; GENERAL INFORMATION:  
; APPLICANT: Nikolau, Basil J  
; APPLICANT: Wurtele, Eve S  
; APPLICANT: Oliver, David J  
; APPLICANT: Behal, Robert  
; APPLICANT: Schnable, Patrick S  
; APPLICANT: Ke, Jinshan  
; APPLICANT: Johnson, Jerry L  
; APPLICANT: Allred, Carolyn C  
; APPLICANT: Fatland, Beth  
; APPLICANT: Lutziger, Isabelle  
; APPLICANT: Wen, Tsui-Jung  
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and  
; FILE OF INVENTION: Acetyl CoA Levels in Plants  
; FILE REFERENCE: P2194USDIV-2  
; CURRENT APPLICATION NUMBER: US/11/167,856  
; CURRENT FILING DATE: 2005-06-27  
; PRIOR APPLICATION NUMBER: US 10/293,865  
; PRIOR FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: US 09/344,882  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 60/090,717  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 3.1  
; SEQ ID NO 20  
; LENGTH: 538  
; TYPE: PRT  
; ORGANISM: Arabidopsis Thaliana  
US-11-167-856-20

Query Match 3.2%; Score 87; DB 7; Length 538;  
Best Local Similarity 27.5%; Pred. No. 2.2; Mismatches 47; Indels 42; Gaps 8;  
Matches 39; Conservative 14;  
QY 148 AKTRFDGKPNWSPSLDRPEAPALAREEELAEVALYEWIQLFWLWGTQKAYAES 207  
Db 103 AARTADEGPWPOMSAVERSVLLRFDLVEKHSSELASLE-----TWDNGKPYQOS 154  
QY 208 KGITIGDMPIFV-AFSDSDVANPOY-FYLEADGN-----PTVWAG--VPRDYFSE 255  
Db 155 ----LTAIEPMFARLFYAGWADKIHGLTIPADGNVQVHTLHEPIGVAGQIIP-----204  
QY 256 TGOIWMGNPLRYWDMERDNEAFW 277  
Db 205 ----WNFPLLM-----PAW 214

## RESULT 9

US-11-012-762-64  
; Sequence 64, Application US/11012762  
; Publication No. US20050244815A1  
; GENERAL INFORMATION:  
; APPLICANT: Georgia State University Research Foundation, Inc.  
; TITLE OF INVENTION: Compositions and Methods for Viral Resistance Genes  
; FILE REFERENCE: GSUI.PCT  
; CURRENT APPLICATION NUMBER: US/11/012,762  
; CURRENT FILING DATE: 2004-12-15  
; PRIOR APPLICATION NUMBER: PCT/US03/19300  
; PRIOR FILING DATE: 2003-06-19  
; CURRENT APPLICATION NUMBER: US 60/390,046

; PRIOR FILING DATE: 2002-06-19  
; NUMBER OF SEQ ID NOS: 133  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 64  
; LENGTH: 512  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-11-012-762-64

Query Match 3.2%; Score 86; DB 7; Length 512;  
Best Local Similarity 18.1%; Pred. No. 2.6; Mismatches 79; Conservative 45; Indels 216; Gaps 21;  
QY 110 LRRAPAGFRARASAOQDKTRLEAFIEAEERFWLEDFALFMALKTRFDGKPNWSPSLDRPE 169  
Db 191 LQSFVVR-----PTKLSLLRLVGVWYQY-----VRDKC 222  
QY 170 PAALARAREELAEVALYEW-----IOMLFYLEWG 199  
Db 223 PRANLPPLVAL-ELLTVYAWAEGTQEDSNFRDLDEGLATVMELLQDHELLCIYW-----274  
QY 200 QTKAYAESKGI-----QIIGMPIFVAFDSSDVANPOYFYLEADGNPTVWAGVPRDY 252  
Db 275 -TKYTTLOHPVIERFVRRLKGERPIIL-----DP-----ADPTNVAQG-----313  
QY 253 FSETGQLWGNPLRYWDMERDNEAFWCIARIQSLKO--CH-----LVRIDHFRG 299  
Db 314 -----YRWDIVAQ-----RASQCLKQDCCYDDRDAPVPSWTVKRAPYIQV 353  
QY 300 FEAYMEVPFGRPNAVEGRWVK--APGEKLFVAVRAQLSDAPIAEDLGVITPEVALRDG 357  
Db 354 TVQOM-----GHPDLI--LWVNPVEPIKL-----KEKIRLS 383  
QY 358 FGPPGMKILQFAGSGEDNAFLPHNPAGNVVYSGTHNDITLGHFRTAPEAEAFMRA 417  
Db 384 RGYSGQLRSLF-----QBPGGQRLIRS 406  
QY 418 Y--LARYGIRCLSE--YEVAGALIELAFKSP-----AKLAIVPLQDVILG-----458  
Db 407 QCSLAYIGIFCDTQICLLDTISPEIOVFVNPDGSHAYAIHPLDFVLSLKQOIEDRQGL 466  
QY 459 -GPEARMNFPGRGDNW 474  
Db 467 QSQEQQLFEQGRVLEDM 483

## RESULT 10

US-11-038-284-34  
; Sequence 34, Application US/11038284  
; Publication No. US20050246793A1  
; GENERAL INFORMATION:  
; APPLICANT: COOKE, DAVID  
; APPLICANT: DEBET, MARTINE  
; APPLICANT: GIDLEY, MICHAEL, JOHN  
; APPLICANT: JOBLING, STEPHEN, ALAN  
; APPLICANT: SAFFORD, RICHARD  
; APPLICANT: SIDEBOTTOM, CHRISTOPHER, MICHAEL  
; APPLICANT: WESTCOTT, ROGER, JOHN  
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO PLANT STARCH COMPOSITION  
; FILE REFERENCE: 054163-5003-US  
; CURRENT APPLICATION NUMBER: US/11/038,284  
; CURRENT FILING DATE: 2005-01-21  
; PRIOR APPLICATION NUMBER: US/10/056,454  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: PCT/GB96/01075  
; PRIOR FILING DATE: 1996-05-03  
; PRIOR APPLICATION NUMBER: GB 9607409.1  
; PRIOR FILING DATE: 1996-04-10  
; PRIOR APPLICATION NUMBER: GB 9509229.2  
; PRIOR FILING DATE: 1995-05-05  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 34

```

; LENGTH: 695
; TYPE: PRT
; ORGANISM: Solanum tuberosum
US-11-038-284-34

Query Match      3.28; Score 85.5; DB 7; Length 695;
Best Local Similarity 18.38; Pred. No. 4.3;
Matches 67; Conservative 40; Mismatches 106; Indels 153; Gaps 18;

QY 54 YGDSPPYQSFAFA-----GNP-----YLVDP-----EMLIE-----KGM-L 83
DB 218 YGSFGYHVTNFFAVSNRYGNPDLKYLDKAHSLGLQVLVDVVHSHASNVTDLGNFDI 277
QY 84 EQSEAPPPYQRTQVDYGLWLYQTR-----WPLLRAPAGFRARASQDKTRLEAFIEAR 137
DB 278 GQSQSSEYFHAGERGYHKLWDSRLFNVANWEVLRLLSNL-----R 318
QY 138 FWEEDYAL--FMALKTRFDG-----KPMNWSPELRDRPAALARAREE 179
DB 319 WMLEEYNFDG-----RFDGITSMLYVHGINMGFTGNYEYFSEATDVDVAVYLMANN 373
QY 180 L-----AEEVALYEWI-----QWLFYL-----EWGQT 201
DB 374 LIHKIFPDATVIAEDVSGMGLSPVSEGGIGFDYRLAMAI PKDWIDYLNKKNDEWMSK 433
QY 202 K-----AYAESKIQIIGDMPIFVAFDSSDVWNPQFYFYLEADGNPTVA 246
DB 434 EVTSSLNRRYTEKCIAYAESHDQSIQVGDKTIAFLMLDKEMYSGMSCL-----TDASPVVDR 490
QY 247 GVP-----RDYFSETGQMGNPLRYRDWMDRDNFAMCIARQSLKQCHLV 292
DB 491 GIALHKMIHFTWALGEGYLNFMNFGFHP--EWIDFPREGNNWSYDKCR-----RWNLA 545
QY 293 RIDHFR 298
DB 546 DSEHLR 551

RESULT 11
US-11-205-109-14
; Sequence 14, Application US/11205109
; Publication No. US20050287641A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
; FILE REFERENCE: 3002-2US
; CURRENT APPLICATION NUMBER: US/11/205,109
; CURRENT FILING DATE: 2005-08-17
; PRIOR APPLICATION NUMBER: US/09/976,059
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/239,924
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 6893
; TYPE: PRT
; ORGANISM: Actinoplanes sp.
US-11-205-109-14

Query Match      3.18; Score 85; DB 7; Length 6893;
Best Local Similarity 19.88; Pred. No. 1e+02;
Matches 101; Conservative 44; Mismatches 153; Indels 212; Gaps 26;

QY 17 PGRWGIGAL-----GREARFLD-WLADAGARWQVPLGPTSYGDSFYQSFSFAGNPLYL 71
DB 1515 PGR-ALGTLDDVAGRRDLDDGNA-----TAVPAQPAL 1548
QY 72 VDPKMLEKHWLEQSEAPPYPTQRTQVDYGLWLYQTRWPLLRAPAGFRARASQDKTRLEA 131
DB 1549 V-PEL-----FTAQAATP-----TWPLVTAAGAEMSYAELEERSNRLAR 1587

```

```

QY 132 FIEAERFWLED-----VALFMALKTRFDGKPMNWSPEL--RDRBPAALARAR 177
DB 1588 WLAGRGVGADDRVALMRRRGPELMVAIILAVLKA---GAAYLPVDPDLPRDRVDYLLADAA 1644
QY 178 EE--LAEEVALYEWIQWLFYLEWGTQKAYAESKIQIIGDMPIFVAFDSSDVWNPQFY 235
DB 1645 PAFVLAER-----ATAPWVPVAGGIPVLVV-----1669
QY 236 LEADGNPTVAGVPRDYSFSETGQMGNPLRYRDWMDRDNFAMCIARQSLKQCHLVRI 295
DB 1670 ----DAPAAAEVA-----AHSGEA-----VTDRDR-----RAALRGHLYAYVI 1704
QY 296 HFRGFEAYWEVFPGRPNAYEGRWKAPGKLEAAVRAQLSDAPIIAEDLGVITPEVEALR 355
DB 1705 YTSG-----STGRPKGV-----LITHD-GLANLTLDHGR 1732
QY 356 DGFPGPKMILQAFSGEDNAPLPHNYPAHNNVYVGGTHDNDTTLGWFRTPAEAR--A 413
DB 1733 FGLG-PGARVAQFASPGFD-----MFVDEWSMALLAGAA-----LTFVPPERRLGA 1777
QY 414 FRAVILARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDDVLGLGPEARMNPPGRLDN 473
DB 1778 DLAFLAEYG-----VTHATLPPAVVGTIPDGLVPPSFVLDVGVDV---LPGDLARR 1826
QY 474 W-----AWRYAEGD 482
DB 1827 WLDRGRVLENSYGPTETTVNAATWRAEAGD 1856

RESULT 12
US-10-454-437-102
; Sequence 102, Application US/10454437
; Publication No. US20050277115A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; INVOLVED IN HOMEOSTASIS AND ADAPTATION
; FILE REFERENCE: BGI-128CPCN
; CURRENT APPLICATION NUMBER: US/10/454,437
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931636.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932126.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932127.2
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932129.9
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: DE 19932226.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932920.6
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932922.2
; PRIOR FILING DATE: 1999-07-14
; Remaining prior application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 442
; SEQ ID NO 102
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-454-437-102

```



Query Match 3.1%; Score 84.5; DB 6; Length 355;  
Best Local Similarity 21.2%; Pred. No. 2.1;  
Matches 68; Conservative 33; Mismatches 87; Indels 133; Gaps 19;  
QY 127 TLEAFIEAEERFWLELYALFMALKTRFDGKPNWNSPELRDREPAALARAR-----177  
DB 68 THADYVIEET--ADVAARAL-----LYDLAPAASANAKEYKLSAEE 110  
QY 178 -BELAEVALYEWIOWLFYLEWGQTKAYAESKGIQIIGDMPFVAFDSSDVWNPQYFYL 236  
DB 111 AEHGOKVGAADWSF-----EAKVRARD-----YALDETEL-----RNYFPL 148  
QY 237 EADGNPTVAVGPRDYFSETGOLWG-----NPLYR-----WDYMERDNFAWCIRAR 283  
DB 149 ----NOVLADGV-----FFAANRLYGITVEPRDLRGYAGVDVWEVLDSDG-----191  
QY 284 QSLKQCHLVRIHFRGFAYWEVPPGRPNVAGRWKAPGEKLFPAVRAQLSDAPIIAD 343  
DB 192 ---SGIGLITDY-----YGRPSKRGGMSS-----FVDQSELLGTPVNVV 232  
QY 344 LGVITPEV-EALRD-----GFGPPGM--KILQAFSG-----ED 374  
DB 233 MGITPTTGEALLSDEVTTFHFEGHGLHLLSKVRYPFSFGTSVPRDYVFPFSQINEN 292  
QY 375 NAFLP---HNYPAH---GNVV 389  
DB 293 WAFDPAVVRYARHVDGTGDI 313

RESULT 13  
US-10-454-437-100  
; Sequence 100, Application US/10454437  
; Publication No. US20050277115A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Haberhauer, Gregor  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS  
; FILE OF INVENTION: INVOLVED IN HOMEOSTASIS AND ADAPTATION  
; FILE REFERENCE: BGI-128CPCN  
; CURRENT APPLICATION NUMBER: US/10/454,437  
; PRIOR FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: DE 19931636.8  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19932125.6  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932126.4  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932127.2  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932128.0  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932129.9  
; PRIOR FILING DATE: 1999-07-19  
; PRIOR APPLICATION NUMBER: DE 19932226.0  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932920.6  
; PRIOR FILING DATE: 1999-07-14  
; PRIOR APPLICATION NUMBER: DE 19932922.2  
; PRIOR FILING DATE: 1999-07-14  
; REMAINING PRIOR APPLICATION data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 442  
; SEQ ID NO 100  
; LENGTH: 555  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-10-454-437-100  
Query Match 3.1%; Score 84.5; DB 6; Length 555;  
Best Local Similarity 21.2%; Pred. No. 2.1;  
Matches 68; Conservative 33; Mismatches 87; Indels 133; Gaps 19;  
QY 127 TLEAFIEAEERFWLELYALFMALKTRFDGKPNWNSPELRDREPAALARAR-----177  
DB 68 THADYVIEET--ADVAARAL-----LYDLAPAASANAKEYKLSAEE 110  
QY 178 -BELAEVALYEWIOWLFYLEWGQTKAYAESKGIQIIGDMPFVAFDSSDVWNPQYFYL 236  
DB 111 AEHGOKVGAADWSF-----EAKVRARD-----YALDETEL-----RNYFPL 148  
QY 237 EADGNPTVAVGPRDYFSETGOLWG-----NPLYR-----WDYMERDNFAWCIRAR 283  
DB 149 ----NOVLADGV-----FFAANRLYGITVEPRDLRGYAGVDVWEVLDSDG-----191  
QY 284 QSLKQCHLVRIHFRGFAYWEVPPGRPNVAGRWKAPGEKLFPAVRAQLSDAPIIAD 343  
DB 192 ---SGIGLITDY-----YGRPSKRGGMSS-----FVDQSELLGTPVNVV 232  
QY 344 LGVITPEV-EALRD-----GFGPPGM--KILQAFSG-----ED 374  
DB 233 MGITPTTGEALLSDEVTTFHFEGHGLHLLSKVRYPFSFGTSVPRDYVFPFSQINEN 292  
QY 375 NAFLP---HNYPAH---GNVV 389  
DB 293 WAFDPAVVRYARHVDGTGDI 313

RESULT 14  
US-10-858-730-22  
; Sequence 22, Application US/10858730  
; Publication No. US20050255568A1  
; GENERAL INFORMATION:  
; APPLICANT: Bailey, Richard B.  
; APPLICANT: Blomquist, Paul  
; APPLICANT: Doten, Reed  
; APPLICANT: Driggers, Edward M.  
; APPLICANT: Madden, Kevin T.  
; APPLICANT: O'Leary, Jessica  
; APPLICANT: O'Toole, George  
; APPLICANT: Trueheart, Joshua  
; APPLICANT: Walbridge, Michael J.  
; APPLICANT: Yorgey, Peter S.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID  
; TITLE OF INVENTION: PRODUCTION  
; FILE REFERENCE: 14184-030001  
; CURRENT APPLICATION NUMBER: US/10/858,730  
; CURRENT FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/475,000  
; PRIOR FILING DATE: 2003-05-30  
; PRIOR APPLICATION NUMBER: US 60/551,860  
; PRIOR FILING DATE: 2004-03-10  
; NUMBER OF SEQ ID NOS: 364  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 379  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-858-730-22  
Query Match 3.1%; Score 84; DB 6; Length 379;  
Best Local Similarity 18.0%; Pred. No. 2.6;  
Matches 93; Conservative 48; Mismatches 139; Indels 236; Gaps 23;  
QY 22 IGALGREAREFLDLADAGARWQVLPGLPTSYGDSYQSFSAFAGNPNVLPDPEMLIEKG 81  
DB 22 VGSQLESQAVIDDVCIQAVQWGLSPA-----RDNVVVLHALGDSHITGP-----69  
QY 82 WLEQSEAPPPYPTORVDYGH-----LYQTFW-PILRFAFAGFRASAQDKTRLE 130  
DB 70 -----AGPGHPTP-----GMWDGVAGPGAPIDTTTRWCATVATNVLGCCRGSGTSPSLAR--117  
QY 131 AFIEAEERFWLELYALFMALKTRFDGKPNWNSPELRDREPAALARAREBEVALYEWI 190

```
Db 118 -----DGPWGRFPLISIRD----- 133
QY 191 QMLFYLEWGQTKAYAESKGIQIICDMPFIFVAFDSSDVWANPOFYLEADGNPTVVAGVPR 250
Db 134 -----QVQADVAALAAALGITEVAA----- 155
QY 251 DYFSETGQALWGNPLYRWDVWERNDFWACIARIQSLKQCHLVRIHDHFRGEAYWEVPPGR 310
Db 156 -----GSMGGAARALEWV----- 170
QY 311 ENAVEGRWVKAPECKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF----- 360
Db 171 PDRVRAGLLAVG----- 217
QY 361 ---PGMKILQ---FAFSGE---DNAFLPHNYPAHGNVVVYSGTHDNDTTLGWERTAPEA 410
Db 218 APDAGLRLARRFAHLTYRGEIELDTRFANEN---QGN-----EDPTAG----- 257
QY 411 ERAFMRAYLARYGIRCLSEYEYVAGALIEL--AFKS-----PAKLAIVLQDVILGL 458
Db 258 GRVAVQSYLEHQDKLLSRFD-AGSYVILTEALNSHDVGRGVSAAALRACEVPPVVG-- 315
QY 459 GPEARWVFCRLGDNWARYAEGDLPEGLAAGLAL 494
Db 316 GITSDRLYPLRLOQELA-----DLLPG-CAGLRVV 344
```

## RESULT 15

```
US-10-467-657-70
; Sequence 70, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 70
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-70
```

```
Query Match 3.0%; Score 82; DB 6; Length 715;
Best Local Similarity 19.8%; Pred. No. 9.3;
Matches 83; Conservative 41; Mismatches 136; Indels 160; Gaps 23;

QY 64 AFAGNPLYNDPEMLIEKGMLEQSEAPPYPTQ-----RVDYGMWLYQT--RWPLLRRA--FA 115
Db 22 AFTGN-----ADGWFDAGEPPAPAAQAEGRFRFDYEPFYNVGHSKPREREAVGFA 71
QY 116 GFRARASAOQKTRLEAFIERFWLEDYALFMALKTRFDGKPNWNSPELDRDREPAALAR 175
Db 72 QLRALADNYDVLRL--VIEARKQME-----CLK-----WTIQKRDVESTEDDESQR 116
QY 176 AREELAEVALYE-----WIQW-----LFYLE-----WGQTKAVAESKGIQ 211
Db 117 KDRKVDVAFAFFSPDKHEHTWADWLRILEDLFDVIDAPCIYPRKTLGGGLYALVEMDGT 176
QY 212 I-----IGDMPFVAFDSSDVWANPOFYLEADGNPTVVAGVPR-DYFSETGQLMGNPL 264
Db 177 IKRVLDTNGRML-----PPDTAYQQ-----ILHGMAADVDTAD-----ELI 213
QY 265 YRWDVMERDNFAM-----CIAIROSLK-QCHLVR-----I 294
```

```
Db 214 YR-----SNNRSYKYGYSPVEQIIMTYNIALKROVHALEYVYTAGSVDPDALVGPETWSA 269
QY 295 DHFRGEAYWEVPPFGRPNAVEGRWVKAPE---KLFAAVRA-----Q 333
Db 270 DDIRRFQBYWDLLLSGETAQREKMRFPVPGELSRNFRRETKOPPLKDVYDEWLARVVCFAFS 329
QY 334 LSDAPIIAEDLGVIT--TPEVEALRDGFG-----FPGMKILOFAPSGEDN 375
Db 330 VEPTFPVAVQVNRSAETSRQSLSDOMGSLKMWKALIDDLVRLARYMDMAAYEFVWKEES 369
```

Search completed: January 10, 2006, 14:25:13  
Job time : 30 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2006, 13:56:47 ; Search time 134 Seconds  
(without alignments)  
1642.752 Million cell updates/sec

Title: US-10-618-976-2  
Perfect score: 2714  
Sequence: 1 MQLQRAFGILLHPTSPFGRW.....DLEPGLAAGLALAEASQRA 501

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*  
9: Geneseq2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2714	100.0	501	AAE01153	Thermus
2	1696.5	62.5	500	AAW83330	Thermus
3	1696.5	62.5	500	ADQ15011	Thermus
4	1692.5	62.4	500	ADQ23307	Protein
5	1367	50.4	496	ADU07729	Amino aci
6	1222.5	45.0	505	ADQ15026	Synechocy
7	1196	44.1	505	ADK47713	Streptoco
8	1194	44.0	505	ABU02622	S. pneumo
9	1108.5	40.8	487	ADQ15028	Clostridi
10	1100	40.5	485	ADQ23309	Protein o
11	1076	39.6	458	ADK94304	Novel S.
12	1076	39.6	458	AEA58174	Streptoco
13	1041	38.4	497	ABP28631	Streptoco
14	1035.5	38.2	576	ADQ15022	Potato am
15	1020	37.6	498	ABP28630	Streptoco
16	1020	37.6	498	ADV88490	Streptoco
17	1020	37.6	498	ADV81901	Streptoco
18	1020	37.6	498	ADV79743	Streptoco
19	894.5	33.0	486	ADK94539	Plant ful
20	777.5	28.6	298	ADG98854	Streptoco
21	694	25.6	291	ADG98855	Potato 4-
22	606.5	22.3	470	ABM90697	M. xanthu
23	451	16.6	526	AAV34929	C. pneumo
24	451	16.6	526	ABB94330	Chlamydia

25	451	16.6	526	8	ADQ15030	Chlamydia
26	430.5	15.9	527	5	ABB94365	Chlamydia
27	383	14.1	583	6	ABU03132	Alpha amy
28	383	14.1	583	8	ADU07478	Amino aci
29	361	13.3	388	2	AAV37126	Protein i
30	356.5	13.1	698	5	AAW49553	Actinopla
31	333	12.3	183	2	AAW83189	Corn 4-al
32	333	12.3	727	6	ABM65372	Propionib
33	328	12.1	183	7	ADG98857	Corn 4-al
34	323.5	11.9	204	8	ADK92724	Plant ful
35	313.5	11.6	684	8	ADS14923	Pseudomon
36	301	11.1	696	6	ABM70485	Phototrab
37	299	11.0	681	7	ABO75883	Pseudomon
38	289	10.6	694	8	ADQ15024	Escherich
39	278	10.2	743	7	ABO64556	Klebsiell
40	273.5	10.1	496	2	AAW83188	Corn 4-al
41	273.5	10.1	496	7	ADG98851	Corn 4-al
42	272	10.0	153	4	AAU03658	Group B S
43	271	10.0	584	4	AAU44938	Propionib
44	271	10.0	584	6	ABM41457	Propionib
45	264	9.7	721	5	ABP66204	Bifidobac

ALIGNMENTS

RESULT 1  
AAE01153  
ID AAE01153 standard; protein; 501 AA.  
XX  
AC AAE01153;  
XX  
DT 06-AUG-2003 (revised)  
DT 17-JUL-2001 (first entry)  
XX  
DE Thermus rubens glucanotransferase protein.  
XX  
KW Glucanotransferase; starch; Japanese dessert; noodles; wheat product;  
KW sea food; processed food; baby; animal; drink; gyoza skin; shumai skin;  
KW nutrient supplemental food; snack; detergent composition.  
XX  
OS Thermus sp.  
XX  
PN WO200129195-A1.  
XX  
PD 26-APR-2001.  
XX  
PF 16-OCT-2000; 2000WO-DK000560.  
XX  
PR 20-OCT-1999; 99DK-00001501.  
PR 15-NOV-1999; 99DK-00001641.  
XX  
PA (NOVO ) NOVOZYMES AS.  
XX  
PI Fukuyama S;  
XX  
DR WPI; 2001-290912/30.  
DR N-PSDB; AAD04867.  
XX  
PT Novel polypeptides having glucanotransferase activity for producing food  
and in detergent or cleaning compositions for removing starch stains,  
especially amylose stains.  
XX  
PS Claim 1; Page 52-54; 57pp; English.  
XX  
CC The present sequence is Thermus rubens ATCC 31556 glucanotransferase.  
CC Glucanotransferase catalyses the cyclisation of starch by acting on  
starch in food to produce a cyclic glucan. Glucanotransferase is useful  
for producing food such as Japanese desserts, snacks, wheat products,  
noodles, gyoza skins, shumai skins, processed sea foods, frozen or  
refrigerated processed foods, weaning foods, baby foods, pet foods,  
animal feeds, drinks, sports foods and nutrient supplemental foods. A  
cleaning or detergent composition comprising glucanotransferase is useful

CC	for removal of starch stains, in particular for removal of amylose	
CC	stains, from a hard surface or from laundry, by contacting the amylose	
CC	stain-containing hard surface or the amylose stain-containing laundry	
CC	with glucanotransferase or its composition. Nucleic acid constructs,	
CC	vectors and host cells comprising glucanotransferase nucleic acid	
CC	sequence are also provided. (Updated on 06-AUG-2003 to correct OS field.)	
XX		
SQ	Sequence 501.AA;	
	Query Match 100.0%; Score 2714; DB 4; Length 501;	
	Best Local Similarity 100.0%; Pred. No. 1.7e-255;	
	Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MOLQAFGILLHPTSPFGRWGIGALGREARFIDLWADAGARWQVLPGLGTSYGDSPYQ 60	
DB	1 MOLQAFGILLHPTSPFGRWGIGALGREARFIDLWADAGARWQVLPGLGTSYGDSPYQ 60	
QY	61 SFSAFAGNPLYVDPEMLIEKGMLEQSEAPPPYPTQRYDYGWLYQTRWPLLRRAAFAGFRAR 120	
DB	61 SFSAFAGNPLYVDPEMLIEKGMLEQSEAPPPYPTQRYDYGWLYQTRWPLLRRAAFAGFRAR 120	
QY	121 ASAQDKTRLEAFTEAEERFWLEDYALFWALKTRFDGKPNWSPELRDREPAALARAREEL 180	
DB	121 ASAQDKTRLEAFTEAEERFWLEDYALFWALKTRFDGKPNWSPELRDREPAALARAREEL 180	
QY	181 AEEVALYEWIOWLFYLEWGOTKAYAESKGIQIIGDMPIFVAFDSSDVWNPQVLYLEADG 240	
DB	181 AEEVALYEWIOWLFYLEWGOTKAYAESKGIQIIGDMPIFVAFDSSDVWNPQVLYLEADG 240	
QY	241 NPTVAGVPDYSSETGOLGNPNLYRWDMVRDNFACIARIQSLSKQCHLVRIIDHPRGF 300	
DB	241 NPTVAGVPDYSSETGOLGNPNLYRWDMVRDNFACIARIQSLSKQCHLVRIIDHPRGF 300	
QY	301 EAYWEVFGPNNAVEGRWVKAPEKLPFAAQAQLSDAPIIAEDLGVTTPVEALRDGFGF 360	
DB	301 EAYWEVFGPNNAVEGRWVKAPEKLPFAAQAQLSDAPIIAEDLGVTTPVEALRDGFGF 360	
QY	361 PGMKILOFASGDNFALPHNYPAGHNVVYSGTHDNDTTLGWFRTPAPEARAFMRAYLA 420	
DB	361 PGMKILOFASGDNFALPHNYPAGHNVVYSGTHDNDTTLGWFRTPAPEARAFMRAYLA 420	
QY	421 RYGRICLSEYEVAGALIELAFKSPAKLAIPLQDVLGLGPEARNMFPGRIGDNWAWRYAE 480	
DB	421 RYGRICLSEYEVAGALIELAFKSPAKLAIPLQDVLGLGPEARNMFPGRIGDNWAWRYAE 480	
QY	481 GDLEPGLAAGLRAALAEASORA 501	
DB	481 GDLEPGLAAGLRAALAEASORA 501	
RESULT 2		
AAW83330		
ID	AAW83330 standard; protein; 500 AA.	
XX		
AC	AAW83330;	
XX		
DT	17-OCT-2003 (revised)	
DT	01-MAR-1999 (first entry)	
XX	Thermus flavus amylomaltase.	
DE		
XX	Thermus flavus; amylomaltase; heat resistant; cyclic glucan;	
KW	intramolecular transglycosylation; alpha-glucan; food.	
KW		
OS	Thermus thermophilus.	
XX		
PN	EP884384-A2.	
XX		
PD	16-DEC-1998.	
XX		
PF	13-MAY-1998; 98EP-00250162.	
XX		
PR	13-MAY-1997; 97JP-00122635.	

PR	07-MAY-1998; 98JP-00125121.	
XX		
PA	(EZAK ) EZAKI GLICO CO LTD.	
XX		
PI	Terada Y, Fujii K, Yanase M, Takata H, Takaha T, Okada S;	
XX		
DR	WPI; 1999-026580/03.	
DR	N-PSDB; AAV72539.	
XX		
PT	New amylomaltase catalyses intra-molecular trans-glycosylation of alpha-	
PT	glucans - used as additive to e.g. rice products, snacks, wheat products,	
PT	noodles, processed seafood, frozen or refrigerated foods, baby foods or	
PT	drinks.	
XX		
PS	Claim 3; Page 17-18; 32pp; English.	
XX		
CC	The present sequence is an amylomaltase isolated from Thermus flavus.	
CC	Amylomaltase catalyses intramolecular transglycosylation of alpha-glucans	
CC	to generate cyclic glucans, has no hydrolase activity, has an optimum	
CC	temperature of 65-70 degrees Celsius, remains active at 60 degrees	
CC	Celsius for at least 10 minutes, is inactivated at 100 degrees Celsius	
CC	within 15 minutes and has an optimum pH of 5.5. Amylomaltase can be used	
CC	to produce a cyclic glucan by cyclising an alpha-glucan and collecting	
CC	and purifying the cyclic glucan (especially where the cyclic glucan	
CC	comprises a cyclic alpha-1,4-glucan, a branched cyclic glucan and/or a	
CC	branching enzyme is also used in the cyclisation step). The amylomaltase	
CC	is used to produce food by adding it to a food material before or	
CC	immediately after cooking so that the amylomaltase acts on starch in the	
CC	food material to produce a cyclic glucan (especially where the food is	
CC	selected from rice products, Japanese desserts, snacks, wheat products,	
CC	noodles, gyozu skins, shumai skins, processed seafoods, frozen or	
CC	refrigerated processed foods, weaning foods, baby foods, pet foods,	
CC	animal feeds, drinks, sports foods and nutritional supplements). (Updated	
CC	on 17-OCT-2003 to standardise OS field)	
XX		
SQ	Sequence 500 AA;	
	Query Match 62.5%; Score 1696.5; DB 2; Length 500;	
	Best Local Similarity 62.9%; Pred. No. 3.6e-156;	
	Matches 315; Conservative 65; Mismatches 118; Indels 3; Gaps 2;	
QY	1 MOLQAFGILLHPTSPFGRWGIGALGREARFIDLWADAGARWQVLPGLGTSYGDSPYQ 60	
DB	1 MELPRAFGILLHPTSLPGYGVGLGREARDFLRFLKEAGRGYQWVLPGLGTSYGDSPYQ 60	
QY	61 SFSAFAGNPLYVDPEMLIEKGMLEQSEAPPPYPTQRYDYGWLYQTRWPLLRRAAFAGFRAR 120	
DB	61 SFSAFAGNPLYDLRPLAERGVYVRLED--PGFPQGRVDYGLLYAWKWPALKEAFRPFKEK 118	
QY	121 ASAQDKTRLEAFTEAEERFWLEDYALFWALKTRFDGKPNWSPELRDREPAALARAREEL 180	
DB	119 ASPEEREAFAAFREARAWLEDYALFWALKGAHGLPWRNWPPLRKEEKALREAKSAL 178	
QY	181 AEEVALYEWIOWLFYLEWGOTKAYAESKGIQIIGDMPIFVAFDSSDVWNPQVLYLEADG 240	
DB	179 AEEVAFHFTQWLFRRQWGALEAALGIRIIGDMPIFVAFDSSAEVWAHPWFHLEEG 238	
QY	241 NPTVAGVPDYSSETGOLGNPNLYRWDMVRDNFACIARIQSLSKQCHLVRIIDHPRGF 300	
DB	239 RPTVAGVPPDYFSETGQRWGNPLYRWDLEREGFQFWIRLEKALELFLVLRIIDHPRGF 298	
QY	301 EAYWEVFGPNNAVEGRWVKAPEKLPFAAQAQLSDAPIIAEDLGVTTPVEALRDGFGF 360	
DB	299 EAYWEIPASCTPAVEGRWVKAPEKLFQKIQEVGFEVPLAEDLGVTTPVEALRDGFGF 358	
QY	361 PGMKILOFAF-SGEDNAFLPHNYPAGHNVVYSGTHDNDTTLGWFRTPAPEARAFMRAYL 419	
DB	359 PGMKVQLQAFDDCGMENPFLPHNYPAGHNVVYSGTHDNDTTLGWYRTATPHEKAFMAYIL 418	
QY	420 RYGRICLSEYEVAGALIELAFKSPAKLAIPLQDVLGLGPEARNMFPGRIGDNWAWRYA 479	
DB	419 ADWGITFREEEVFWALMHLGMSKSVARLAVYVQDVLALGSEARNMFPGRPSGNWAWRL 478	

QY 480 EGDLEPGLAAGRLAAASOR 500  
 Db 479 PGELSPHGRLAAEAATER 499

## RESULT 3

ID ADQ15011 standard; protein; 500 AA.

XX AC ADQ15011;

XX DT 23-SEP-2004 (first entry)

XX DE Thermus aquaticus amylo maltase polypeptide.

XX KW Amylo maltase; transgenic; foodstuff; food additive; modifier;  
 KW saccharide; (alpha)-1,4-glucan; cooked rice; Japanese confectionery;  
 KW snack; noodle; Chinese dumpling; shao-mai; fishery kneading;  
 KW refrigerated; baby food; pet; animal feed; drink; food supplement;  
 KW cyclic glucan; Thermus aquaticus; enzyme.

XX OS Thermus aquaticus.

XX PN JP2004187674-A.

XX PD 08-JUL-2004.

XX PF 12-SEP-2003; 2003JP-00322319.

XX PR 26-NOV-2002; 2002JP-00342966.

XX PA (EZAKI) EZAKI GLICO CO LTD.

XX PA (NIDE) NEC CORP.

XX DR WPI; 2004-528719/51.

XX DR N-PSDB; ADQ15010.

XX PT New mutated amylo maltase polypeptide having increased enzyme activity  
 PT and reduced hydrolyzing activity, useful for manufacturing foodstuffs  
 PT such as Japanese confectionery, noodles, baby foods, and food additives.

XX PS Claim 5; SEQ ID NO 2; 51pp; Japanese.

XX CC The invention relates to a novel amylo maltase polypeptide comprising the  
 CC amino acid sequence of a wild-type amylo maltase with a substitution,  
 CC addition or deletion at a position, where the amino acid residue  
 CC interacts with acarbose of amino acids other than the wild-type amylo  
 CC maltase amino acid sequence. The invention further comprises: a nucleic  
 CC acid molecule containing a nucleic acid sequence encoding the novel amylo  
 CC maltase polypeptide; a vector containing the nucleic acid sequence; a  
 CC cell containing the nucleic acid sequence; biological tissue containing  
 CC the nucleic acid sequence; a transgenic organism containing the nucleic  
 CC acid sequence; a foodstuff, a food additive or a modifier of foodstuff  
 CC containing the nucleic acid sequence; a computer readable recording  
 CC medium which contains the information of the nucleic acid sequence  
 CC encoding the amino acid sequence of the novel amylo maltase polypeptide;  
 CC and a cyclic glucan obtained by reacting the novel amylo maltase  
 CC polypeptide on saccharides which have a linear structure of (alpha)-1,4-  
 CC glucan. The novel amylo maltase polypeptide is useful for manufacturing  
 CC foodstuffs such as cooked rice, Japanese confectionery, snacks, bakeries,  
 CC noodles, Chinese dumpling, shao-mai, fishery kneading goods,  
 CC refrigerating foods, baby foods, pet foods, feed for animals, drinks, and  
 CC food supplements, where the method involves adding the novel amylo  
 CC maltase polypeptide to the foodstuff before or immediately after heat  
 CC processing of the raw material, where the novel amylo maltase polypeptide  
 CC generates a cyclic glucan from the starch of the foodstuff. The novel  
 CC amylo maltase polypeptide is useful for manufacturing glucan which has a  
 CC cyclic structure of (alpha)-1,4-glycoside linkage, foodstuff materials  
 CC and food additives, modifier of foodstuffs, a drink or eatable  
 CC composition, infusion solution or composition for adhesion, where the  
 CC method involves reacting the saccharide containing the linear structure  
 CC of (alpha)-1,4-glycoside linkage or its derivatives with the novel amylo  
 CC maltase polypeptide. This sequence represents the Thermus aquaticus amylo

CC maltase polypeptide of the invention.

XX Sequence 500 AA;

Query Match 62.5%; Score 1696.5; DB 8; Length 500;

Best Local Similarity 62.9%; Pred. No. 3.6e-156;

Matches 315; Conservative 65; Mismatches 118; Indels 3; Gaps 2;

QY 1 MOLRAFGLLHPTSPGRWGIGALGREAFDMLADAGARWQVLPGLPTSYGDSPYQ 60

Db 1 MELPRAPGLLHPTSLPGYGVGLGREANDFLRLKEAGRGYQVLPGLPTGYGDSPYQ 60

QY 61 SFSAFAGNPLYVDPEMLIEKQLEQSEAPPYPTQRTVDYGLWYQTRWPLRLRRFAFGRAR 120

Db 61 SFSAFAGNPLYDLRLPLAERGVYRLD--PGFPQGRVDYGLLYANKWPALKAEAFRGFKEX 118

QY 121 ASAQDKTRLEAFIRAFRWLELYALFMALKTRPDGKPNWSEPELDRDRPAALAREEL 180

Db 119 ASPEEREAFAAFREAREAWLEDYALFMALKGAHGLPWRWPLPLRKREKALREAKSAL 178

QY 181 ABEVALYEWIOWLFYLEWGTQKAYAESKGIQIIGDMPIFVAPDSSDVWNPQVYLEADG 240

Db 179 ABEVAFHAFQWLFQWGLKAEALGIRIIGDMPIFVAEDSARVWAHPWFHDEEG 238

QY 241 NPTVAGVPRDYFSETGQLWGNPLYRWDMERNFAWCIAIRQSLKQCHLVRIIDHFRGF 300

Db 239 RPTVAGVPPDYFSETGQWGNPLYRWDLEREGFSFWIRLEKALELFLHVLVIDHFRGF 298

QY 301 EAYWEVFPORPNAVEGRWVKAPGEKLFAAVRAQLSDAPIAEDLGVIITPEVEALRDGFGF 360

Db 299 EAYWEIPASCPATAVEGRWVKAPGEKLFQKIQEVGEVPPVLAEDLGVIITPEVEALRDGFGF 358

QY 361 PGWKILQFAF-SGEDNAFLPHNYPAHGNVNVYSGTHDNDTTLGWFTAPAEAEAFMAYIL 419

Db 359 PGWKVLPQAFDDGDMENPLPHNYPAHGRVYVYTGTHDNDTTLGWYRTATPHEKAFMAYIL 418

QY 420 ARYGRICLSEYEVAGALIELAFKSPAKLAIVLPQDLVGLQPEARMNFPGLRGDNWARYA 479

Db 419 ADWGITFREEEVPPWALMHLGKSVARLAVYPQDVIALGSEARMNYFGRPSGNWAWRL 478

QY 480 EGDLEPGLAAGRLAAASOR 500

Db 479 PGELSPHGRLAAEAATER 499

RESULT 4

ADQ23307

ID ADC23307 standard; protein; 500 AA.

XX AC ADC23307;

XX DT 18-DEC-2003 (first entry)

XX DE Protein of Thermus thermophilus AMase.

XX KW recombinant; alpha-glucanotransferase activity; hydrolysing; hydrolytic;  
 KW alphaalpha/betabeta; barrel structure; transferase; bakery product; bread;  
 KW amyloamaltase; branching enzyme; thermostable transferase;  
 KW thermophilic microorganism; EC 2.4.1.25; EC 2.4.1.18;  
 KW starch retrogradation; amylopectin; alpha-amylase; AMase.

XX OS Thermus thermophilus.

XX PN WO2003002728-A2.

XX PD 09-JAN-2003.

XX PF 01-JUL-2002; 2002WO-NL000427.

XX PR 29-JUN-2001; 2001NL-01018426.

XX PA (NEDE) NEDERLANDSE ORG TOEGEPAST.

DB 419 ADWGITFREEEVPMALHGMKSVARLAVYPQDVLALGSEARMNYPGRPSGNWAWRL 478

QY 480 EGDLEPLAAGRALAASQR 500

DB 479 PGELSPEHGARLMAEAATER 499

RESULT 5

ADU07729

ID ADU07729 standard; protein; 496 AA.

XX AC ADU07729;

XX DT 13-JAN-2005 (first entry)

DE Amino acid sequence for amylase #171.

XX AMylase; glucoamylase; 1,4-alpha-D-glucan glycohydrolase; alpha-amylase;

KW exoamylase; beta-amylase; glucosidic bond hydrolysis; starch; sugar;

KW glucose; maltodextrin; thermostable; enzyme.

XX Unidentified.

OS WO2004091544-A2.

PN 28-OCT-2004.

XX 08-MAR-2004; 2004WO-US0007096.

PF 06-MAR-2003; 2003US-00385305.

XX 28-MAR-2003; 2003US-0459014P.

PR (DIVE-) DIVERSA CORP.

PA Callen W, Richardson T, Frey G, Gray K, Kerovuo JS, Slupska M;

XX Barton N, O'donoghue E, Miller C;

PI WPI; 2004-775532/76.

XX N-PSDB; ADU07728.

DR Novel recombinant amylases, useful for hydrolysis of starch into sugars,

XX in detergent composition, for producing high-maltose or high-glucose

PT syrup, in brewing or alcohol production, bio-bleaching composition of

PT paper or pulp product.

XX Claim 55; SEQ ID NO 439; 436pp; English.

PS The invention relates to the isolation of polypeptides with amylase

CC activity, and polynucleotide sequences encoding the polypeptides. The

CC amylase activity comprises a glucoamylase activity, 1,4-alpha-D-glucan

CC glycohydrolase activity, alpha-amylase activity, exoamylase activity or

CC beta-amylase activity. The amylase activity comprises hydrolyzing

CC glucosidic bonds in a starch to produce sugars such as glucose and

CC maltodextrins. The amylase activity comprises cleaving a maltose or a D-

CC glucose unit from the non-reducing end of the starch. The amylase

CC activity is thermostable and the polypeptides of the invention are active

CC over a wide range of temperatures. The polypeptide sequences are useful

CC for making anti-amylase antibodies, for identifying a modulator of an

CC amylase activity, or identifying an amylase substrate. The polynucleotide

CC sequences are useful for producing recombinant polypeptides, and for

CC modifying codons in a nucleic acid encoding a polypeptide with an amylase

CC activity to increase or decrease its expression in a host cell. The

CC polypeptide and polynucleotide sequences are useful for making or

CC modifying a small molecule, and for determining a functional fragment of

CC an amylase enzyme. Primers to the polynucleotides are useful for

CC amplifying them. This sequence represents a polypeptide sequence of the

XX invention.

CC Sequence 496 AA;

SQ

PI Van Der Maarel MJEC, Dijkhuizen L, Binnema DJ, Van Der Veen BA;

PI Vos A;

XX WPI; 2003-201501/19.

DR N-PSDB; ADC23306.

XX Novel isolated or recombinant nucleic acid encoding a polypeptide with

PT alpha glucanotransferase activity, useful for reducing the retrogradation

PT of starch, amylopectin, and for hydrolyzing starch.

XX Disclosure; Page 79; 92pp; English.

XX The invention relates to a novel isolated or recombinant nucleic acid

CC derived from a nucleic acid encoding a polypeptide essentially having

CC alpha-glucanotransferase activity but having essentially no hydrolysing

CC activity, and the recombinant nucleic acid encoding a polypeptide with

CC hydrolytic activity. The invention further comprises a vector comprising

CC the recombinant nucleic acid; a host cell comprising the vector or the

CC recombinant nucleic acid. The invention also provides a polypeptide or

CC its fragment essentially having alpha-glucanotransferase activity and

CC encoding such a transferase with a mutation leading to an alteration in

CC loss of a codon originally encoding a hydrophobic amino acid located in

CC or around an acceptor, a donor or a catalytic site extending from a

CC (alpha)phalpa/betabeta barrel structure of the transferase. The invention

CC also relates to a bakery ingredient comprising the polypeptide with alpha

CC -glucanotransferase activity and a bakery product such as bread,

CC comprising the polypeptide. The transferase encoded by the recombinant

CC nucleic acid comprises amyloamylase, a branching enzyme or thermotomase

CC transferase. The transferase is derived from a thermophilic microorganism

CC e.g. Thermus thermophilus, T.aquaticus or Aquifex aeolicus. The

CC transferase with enzyme commission (EC) number 2.4.1.25 or 2.4.1.18 is

CC useful in reducing retrogradation of starch and amylopectin, preferably

CC in reducing long-term retrogradation of amylopectin. This transferase is

CC also useful for hydrolysing starch. This sequence represents a Thermus

CC thermophilus AMase protein of the invention.

XX Sequence 500 AA;

SQ

Query Match 62.4%; Score 1692.5; DB 7; Length 500;

Best Local Similarity 62.7%; Pred. No. 8.9e-156;

Matches 314; Conservative 66; Mismatches 118; Indels 3; Gaps 2;

QY 1 MOLORAFGIILHPTSPFGRWGIGALGREAERFLDWLADAGARWQVLPGLGTSVGDSPYQ 60

DB 1 MELPRAFGLLHPTSLPGPYGVGLQGEARDFLFLKEAGRGYQWVLPGLGTSVGDSPYQ 60

QY 61 SFSFAGNPVLVDPEMLIEKGNLQSEAPPYPTQRYDYGLWYOTRWPLLRAPFAGPRAR 120

DB 61 SFSFAGNPVLVDPEMLIEKGNLQSEAPPYPTQRYDYGLWYOTRWPLLRAPFAGPRAR 118

QY 121 ASADKTRLEAFIEAERFLEWEDYALFMALKTRFDGKPNWNSPELRDEPAALAREEL 180

DB 119 ASPEREAFAFREAREAWLEDYALFMALKAHGLGPNRWPLLRKREKALREKASAL 178

QY 181 ABEVALYEWIOWLPYLEWGOTKAYAESKGIQIIGMPIFVAFDSDVWNPQYVLEADG 240

DB 179 AEEVAFHAFTQWLPFRQWALKAEALGIRIIGMPIFVAFDSDVWNPQYVLEADG 238

QY 241 NPTVAGVPRDYFSETGOLGNPLRYDWMERDNFANCIARIOSLKOCHLVRIIDHPRGF 300

DB 239 RPTVAGVPPDYFSETGOLGNPLRYDWMERDNFANCIARIOSLKOCHLVRIIDHPRGF 298

QY 301 EAYVEVPGRENAVEGRWVKAPEKLFVAVRAQLSDAPIIAEDLGVIITPEVEALRDGFGF 360

DB 299 EAYVEIPASCFTAVEGRWVKAPEKLFQKIQEIVGEVPEVLEADLGVIITPEVEALRDGFGF 358

QY 361 PGMKILOFAP-SGEDNAPLPHNYPAHGVVYVSGTHDNDTLGWRTPAPEAFMAYL 419

DB 359 PGMKVLOFAPFDGDMENPPLPHNYPAHGVVYVSGTHDNDTLGWRTPAPEAFMAYL 418

QY 420 ARYGIRCLUSEVEVAGALIELAFKSPAKLAIPLQDVLGLGPEARNPFGRLGDNDWARYA 479

Query Match 50.4%; Score 1367; DB 8; Length 496;

Best Local Similarity 51.9%; Pred. No. 5.3e-124;

Matches 260; Conservative 77; Mismatches 150; Indels 14; Gaps 5;





	Query Match	44.1%	Score 1196	DB 8	Length 505	
Best Local Similarity	48.1%	Pred. No. 2.7e-107				
Matches 242	Conservative 77	Mismatches 168	Indels 16	Gaps 9		
Qy	4	QRAFGILLHPTSPGRWGIGALGREARFLDMLADAGARWQVPLPGTSTGYGDSPIYQSFS	63			
Db	3	KRQSGVLMHTSSLPGAYGIFGSGQSYDFVFLVTRTKQRYWQILPLGATSYGDSPIYQSFS	62			
Qy	64	AFAGNPLYVDPEMLIEKGMLEQSEAPP---PYPTQRVDYGLWLYQTRWPLLRAPAGPRAR	120			
Db	63	AFAGNTHFDLIDLVGGLLEASDLGVDFGSDASEVDYAKIYYARRPLEKAKVRFEV	122			
Qy	121	ASAQDKTRLEAFTEARFMLEDYALPMALKTRFDGKPMNEW-SPELRDREPAALARABEE	179			
Db	123	GDVKD---FEKFAQDNQSMLELFAEYTAIKHEFDNLAWTEWPDADARAKASALESYREQ	179			
Qy	180	LAEEVALYEWIOWLFYLEMGQTKAYAESKGIOIGIMPIFFVAFDSSDVWVNPQVYFLAD	239			



Db 61 FIDILNEKLLDKD-----YEKNFGDNKEMINYGLIFNEKMKVLRKAYMNF-----NS 112  
 Qy 124 QDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPNWSPDLRDPALAREELAE 183  
 Db 113 KDDSPAKFIEDEKDLDDYSLFWALKYKFNFSWNSWKNIDIKLRKNEEIEKYKDELKED 172  
 Qy 184 VALYEWLQWLFYEWGOTKAYAESKGIOLIGDMPIFVAFSSDVTWNPQVYFLEADG-NP 242  
 Db 173 VNYWKFLQYEFFSQQKNLKYANKKNKIKIGDIPYIAQDSSDQVMSNPDIPLLNKETLEP 232  
 Qy 243 TVVAGVPRDYFSETGOLWGNPLRYDWMERDNFANCIARQSLKQCHLVRIIDHFRGFEA 302  
 Db 233 LKWSGCPDPAFSETGOLWGNPIYDMGVLEKTNFEMWDRKSSKLYDLRIDHFRGFEA 292  
 Qy 303 YWEVFGPRNAVEGRWVKAPEKLFPAVRAQLSDAPIAEDLGVTPEVALRDGFGFGP 362  
 Db 293 YWSVDYGEKTAQNGKWKIPKEMKLFNVIKIKIGDIEIIAEDLGVLTEBTELEFKKRTGFG 352  
 Qy 363 MKILOFAPSGE-DNAFLPHNYPAGNVVYSGTHDNDTTLGWPR---TAPAEARAFMAY 418  
 Db 353 MKIIQFAFGDSSNPYLPHNY--EKNCVAYTGTHTDNDTVRGWFVETGSKKEKA----- 405  
 Qy 419 LARYGIRCLSEYEVAGALIELAFKSPAKLATVPLQDVLGLGPEARMNPPGRLGDNMAWR 477  
 Db 406 -VEY-FKLTEEGYNGVIRGWSVANTSICVMQDFNLNGNEARINKPSTLASNWSR 462

## RESULT 10

AD23309  
 ID AD23309 standard; protein; 485 AA.

XX AC AD23309;

XX DT 18-DEC-2003 (first entry)

XX DE Protein of Aquifex aeolicus MTase.

XX KW recombinant; alpha-glucanotransferase activity; hydrolysing; hydrolytic;  
 KW alpha/beta/beta; barrel structure; transferase; bakery product; bread;  
 KW amyloamylase; branching enzyme; thermostable transferase;  
 KW thermophilic microorganism; EC 2.4.1.25; EC 2.4.1.18;  
 KW starch retrogradation; amylopectin; alpha-amylase; MTase.

XX OS Aquifex aeolicus.

XX PN WO2003002728-A2.

XX PD 09-JAN-2003.

XX PF 01-JUL-2002; 2002WO-NL000427.

XX PR 29-JUN-2001; 2001NL-01018426.

XX PA (NEDE ) NEDERLANDSE ORG TOEGEPAST.

XX PI Van Der Maarel MJEC, Dijkhuizen L, Binnema DJ, Van Der Veen BA;  
 Vos A;

XX DR WPI; 2003-201501/19.

XX DR N-PSDB; AD23308.

XX PT Novel isolated or recombinant nucleic acid encoding a polypeptide with  
 PT alpha glucanotransferase activity, useful for reducing the retrogradation  
 PT of starch, amylopectin, and for hydrolysing starch.

XX PS Disclosure; Page 81; 92pp; English.

XX CC The invention relates to a novel isolated or recombinant nucleic acid  
 CC derived from a nucleic acid encoding a polypeptide essentially having  
 CC alpha-glucanotransferase activity but having essentially no hydrolysing  
 CC activity, and the recombinant nucleic acid encoding a polypeptide with  
 CC hydrolytic activity. The invention further comprises a vector comprising  
 CC the recombinant nucleic acid; a host cell comprising the vector or the

CC recombinant nucleic acid. The invention also provides a polypeptide or  
 CC its fragment essentially having alpha-glucanotransferase activity and  
 CC with or without hydrolysing activity, by providing a nucleic acid  
 CC encoding such a transferase with a mutation leading to an alteration or  
 CC loss of a codon originally encoding a hydrophobic amino acid located in  
 CC or around an acceptor, a donor or a catalytic site extending from a  
 CC (alpha/beta/beta) barrel structure of the transferase. The invention  
 CC also relates to a bakery ingredient comprising the polypeptide with alpha  
 CC -glucanotransferase activity, and a bakery product such as bread,  
 CC comprising the polypeptide. The transferase encoded by the recombinant  
 CC nucleic acid comprises amyloamylase, a branching enzyme or thermostable  
 CC transferase. The transferase is derived from a thermophilic microorganism,  
 CC e.g. *Thermus thermophilus*, *T. aquaticus* or *Aquifex aeolicus*. The  
 CC transferase with enzyme commission (EC) number 2.4.1.25 or 2.4.1.18 is  
 CC useful in reducing retrogradation of starch and amylopectin, preferably  
 CC in reducing long-term retrogradation of amylopectin. This transferase is  
 CC also useful for hydrolysing starch. This sequence represents an Aquifex  
 CC aeolicus MTase protein of the invention.

XX SQ Sequence 485 AA;

Query Match 40.5%; Score 1100; DB 7; Length 485;  
 Best Local Similarity 45.8%; Pred. No. 6e-98;  
 Matches 219; Conservative 74; Mismatches 169; Indels 16; Gaps 6;

Qy 5 RAFGILLHPTSPRGRWGIGALGREAFRLDADAGARWQVLPGLPTSY--GDSPYQSF 62

Db 2 RLAGILLHPTSLSPYIGDLGKEAYRFLDLKECGFSLQVLPGLPTSLSEAGNSPYSSN 61

Qy 63 SAFAGNPLYVDPEMLIEKGLQESEAPPYPTQVDYGMVLYOTRPLLRFAFRARAS 122

Db 62 SLFAGNPLYVDPEMLIEKGLQESEAPPYPTQVDYGMVLYOTRPLLRFAFRARAS 117

Qy 123 AQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPNWSPDLRDPALAREELAE 182

Db 118 ---FELLEDLFKEHSYLRDLYALYMAKEE-EGKEWYEWDEELKREKRLKRVLNKLG 173

Qy 183 EVALYEWIOWLFYEWGOTKAYAESKGIOLIGDMPIFVAFSSDVTWNPQVYFLEADGPN 242

Db 174 RYFHFVFOVFFQWKEKRLRYAREGISIVGDLPMYPSYSSADVNTNPELFLKLDGDKP 233

Qy 243 TVVAGVPRDYFSETGOLWGNPLRYDWMERDNFANCIARQSLKQCHLVRIIDHFRGFEA 302

Db 234 LFVAGVPPDFFSKTGOLWGNPNVYNWEEHEKEGFRWIRRVHNLKLFDFLRDLHFRGFEA 293

Qy 303 YWEVFGPRNAVEGRWVKAPEKLFPAVRAQLSDAPIAEDLGVTPEVALRDGFGFGP 362

Db 294 YWEVPGYGEETA VNGRWVKAPEKLTLPKLLSYFPKPNPIAEDLGFTIDEVYRLRETFKIPG 353

Qy 363 MKILOFAPSGEDNAFLPHNYPAGNVVYSGTHDNDTTLGWPRTPAEARAFMAYLARY 422

Db 354 SRVIEFAFYDKSEHLPN--VEENNVYITSTHDLPIRGWENFENLGEESKRLFEYLGRE 411

Qy 423 GIRCLSEYEVAGALIELAFKSPAKLATVPLQDVLGLGPEARMNPPGRLGDNMAWRYAE 480

Db 412 ----IKKEKNEELIRLVLSRAKFAIQMDLLNLGNEARMNYPGPRFGNWRWRIKE 465

## RESULT 11

ADR94304

ID ADR94304 standard; protein; 458 AA.

XX AC ADR94304;

XX DT 16-DEC-2004 (first entry)

XX DE Novel S. pneumoniae protein sequence, SEQ ID 2939.

XX KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;  
 KW bacterial infection.

XX OS *Streptococcus pneumoniae*.



CC	composition and methods are useful for diagnosing, preventing or treating
CC	bacterial infections, particularly S. pneumoniae infection. The present
CC	sequence represents a S. pneumoniae ORF amino acid sequence from the
CC	present invention. Note - The sequence data for this patent did not form
CC	part of the printed specification, but was obtained in electronic format
CC	directly from the USPTO web site.
XX	
SQ	Sequence 458 AA;
Query Match	39.6%; Score 1076; DB 9; Length 458;
Best Local Similarity	48.4%; Pred. No. 1.2e-95;
Matches 221; Conservative	66; Mismatches 154; Indels 16; Gaps 9;
QY	50 GPTSYGSPYQSSAFAGNPIYVDPEMLIEKGMLEQSEAPP---PPTQRYDYGWLYQYR 106
DB	2 GATSYGSPYQSSAFAGNTHFDLIDLVLEQGLLEASDEGVDFGSDASEVDYAKIYYAR 61
QY	107 WPLLRAFAFRARASAOQKTRLEAFTEAERFWLEDYALFMALKTRFDGKPMNEW-SPEL 165
DB	62 RPLEKAVKRFVEVDVKD---FEKPAQDNQSMLELFAEYMAKEHFDNLANTENPDADA 118
QY	166 RDREPAALARABELAEVALYEWIQLFWLWGTQKAYAESKGIQIIGDMPIFVAFDSS 225
DB	119 RAKASALESYRQLADKLVYHRTQYFFQOMLKLYANDNHNIEIVGDMPIYVAEDSS 178
QY	226 DVANPQYFYLEADQNTVTVAGVPRDYFSTGTQGLWGNLYRWDVMERDNFANCIARQS 285
DB	179 DMWANPHLFTDNGKATCIAGCPDDEFSTGTQGLWGNPIYDWEAMDKGYKMWIERLRS 238
QY	286 LKQCHLVIRIDHFRGFAYWEVPGPRNAVGRVWKAPEKLFPAVRAQLSDAPIAEDLG 345
DB	239 FKYDIYRVIDHFRGFESYWEIPAGSDTAAPGEWKGPKLFAAVKEELGELMILIAEDLG 298
QY	346 VITPEVALRDGFGFGMKILQAFSGEDNAP-LPHNYPAGHNVVYVSGTHDNDITLGNF 404
DB	299 FMTDEVELERTGFGFGMKILQAFNPEDESIDSPLAPA--NSVMYTGTHDNDITLGNF 356
QY	405 RTAPEAERAFMAYLARYGIRCLSEYE-VAGALIELAFKSPAKLAIIVLPQDVGLGPEAR 463
DB	357 RN--EIDDA-TREYMARYTNR--KEYETVVMHMLRTVFSSVFMATIMODLLEDEAR 411
QY	464 MNPPGRLGDWAWRYAEGDLEPLGACGLALAEASOR 500
DB	412 MNPFSTLGGNSWRMTEDQLTAPVEEGLDITTIYRR 448
RESULT 13	
ABP28631	standard; protein; 497 AA.
AC	ABP28631;
AC	ABP28631;
DT	02-JUL-2002 (first entry)
XX	Streptococcus polypeptide SEQ ID NO 6438.
XX	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX	group A streptococcus; Streptococcus pyogenes; antibacterial;
XX	antiinflammatory; infection; vaccine; meningitis; gene therapy.
OS	Streptococcus pyogenes.
XX	WO200234771-A2.
XX	02-MAY-2002.
XX	29-OCT-2001; 2001WO-GB004789.
XX	27-OCT-2000; 2000GB-00026333.
XX	24-NOV-2000; 2000GB-00028727.
XX	07-MAR-2001; 2001GB-00005640.
XX	(CHIR-) CHIRON SPA.
PA	(GENO-) INST GENOMIC RES.
XX	Telford J, Masignani V, Margarit Y Rosl, Grandi G, Fraser C;
PI	Tettelin H;
XX	WPI: 2002-352536/38.
DR	N-PSDB; ABN69262.
XX	New Streptococcus protein for the treatment or prevention of infection or
PT	disease caused by Streptococcus bacteria, such as meningitis, and for
PT	detecting a compound that binds to the protein.
XX	Claim 1; Page 3806; 4525pp; English.
XX	The invention relates to a protein (ABP25413-ABP30895) from group B
CC	streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC	(Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC	the specification. The proteins have antibacterial and antiinflammatory
CC	activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC	antibodies that bind (I) are used in the manufacture of medicaments for
CC	the treatment or prevention of infection or disease caused by
CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC	Nucleic acids encoding (I) are used to detect Streptococcus in a
CC	biological sample. (I) is used to determine whether a compound binds to
CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC	used as a vaccine or diagnostic composition. The disease caused by
CC	Streptococcus that is prevented or treated may be meningitis. Nucleic
CC	acid encoding (I) may be used to recombinantly produce (I) and may be
CC	used in gene therapy. Antibodies to (I) are used for affinity
CC	chromatography, immunoassays, and distinguishing/identifying
CC	Streptococcus proteins
XX	
SQ	Sequence 497 AA;
Query Match	38.4%; Score 1041; DB 5; Length 497;
Best Local Similarity	43.6%; Pred. No. 3.6e-92;
Matches 212; Conservative	81; Mismatches 177; Indels 16; Gaps 8;
QY	4 QRAGTLLHPTFFPGRWGIGALGREAFELDWLADAGARWQVLPGLPTSYGSPYQSSFS 63
DB	3 KRASGILMHISLPGKFGIGTGFSAFVDFLAETKQTYWQLPLPTTTSFGSPYQSSFS 62
QY	64 AFAGNPLYVDPEMLIEKGMLEQSE-APPYPT--QRVDYGMWLYQYRPLLRRAFAFRAR 120
DB	63 ATAGNTHFDIDFELLVDLELLEADLCITFGTNPEADYQALFQVRRPLEKAVRAF--V 120
QY	121 ASAQDKTRLEAFTEAERFWLEDYALFMALKTRFDGKPMNEWSE-LRDREPAALARAREE 179
DB	121 AEQENVCKLEAF-ETASSWLTDFAEFMALKEVFNNKALQDWDDETVIKQEDSLNNYREL 179
QY	180 LAEEVALYEWIQLFWLWGTQKAYAESKGIQIIGDMPIFVAFDSSDVANPQYFYLEAD 239
DB	180 LARKITTHKVCQYFFYQOVSALKTYANHKGIEIIGDMPYIVSADSVEVWTPMLFVKVDS 239
QY	240 GNPTVAVGPRDYFSTGTQGLWGNLYRWDVMERDNFANCIARISQLKQCHLVIRIDHFRG 299
DB	240 KXPLFATGVPADGSEDDGQGLWGNPTYNNSAHEKSNFAWMIYRIQESFKLYQLDRIDHFKG 299
QY	300 FEAYWEVFPGRPNVAGRWKAPGEKLFPAVRAQLSDAPIAEDLGIVITEVALRDGFG 359
DB	300 FSDFEIIPAGDKTARNHGWASAPGIALFSAVRALGELPIIAENLGYIDSKAQQLASTG 359
QY	360 FPGMKILOPAFSG--EDNAPLPHNYPAGHNVVYVSGTHDNDITLGFRTAPEAERAFMRA 417
DB	360 FPGMKILEFGLDITTSQSIDLPHYDR--NCVAYTGTNDNEVNGVWTDNLSSEGVHFN 417
QY	418 YLARYGIRCLSEYEYVAGALIELAFKSPAKLAIIVLPQDVGLGPEARNFPRGLGDWAWR 477
DB	418 YLHKH-----ADESITKAMLRITFASVCDTALICIQDLDKDGKSRNMNPTIGGNQWR 472
QY	478 YAEGLD 483
DB	473 MLDGEL 478



Search completed: January 10, 2006, 14:09:43  
Job time : 137 secs